IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,
 	
        Plaintiffs,
 	
    v.
 	
RICK PERRY, *et al.*,
 	
        Defendants.
	Civil Actions No. 2:13-cv-193 (NGR)

DECLARATION OF STEPHEN D. ANSOLABEHERE

Pursuant to 28 U.S.C. § 1746, I, Stephen D. Ansolabehere, make the following

declaration:

HIGHLY CONFIDENTIAL

Supplemental Report

Stephen Ansolabehere

August 15, 2014

I. Statement of Inquiry

1. I have been asked to identify which registered voters in the State of Texas have an acceptable photographic identification required by Texas Senate Bill 14 (2011) (SB 14) and which do not.

2.   I was also asked to determine whether there is a disparity between the percent of Anglo registered voters and Black or Hispanic registered voters who possess acceptable SB 14 photo ID.

3. In order to determine the number of Texas registered voters without acceptable SB 14 photo ID, and to determine any racial disparities in rates of acceptable SB 14 ID between Blacks, Hispanics, and Anglos, I was retained to match records in the Texas Election Administration Management system (TEAM) voter registration database to relevant State of Texas and Federal identification databases.

4. I was provided with data from the TEAM and Texas Department of Public Safety (DPS) databases in February 2014.   The extraction date of these data is January 15, 2014.   In July, the State of Texas determined that its prior production of DPS records was incomplete.   Specifically, Texas did not produce more than 3.1 million DPS records for driver licenses and personal identification cards that should have been included in the original data production.   Upon receiving these additional 3.1 million records, I conducted the analyses in my original report with these additional data. The inclusion of

these additional data lowered the number of records on TEAM for which NO MATCH

could be found on a State of Texas or federal identification database and for which there

was no indication on the DPS database that the individual was deceased from 1.2 million

to 786,727 persons.[1] This Supplemental Report incorporates the additional information

in the record matching process and analyses of the resulting data.

5. I analyzed rates of ID possession for registered voters by race in two different ways.

This is necessary because the TEAM voter registration database does not contain self-

reported information on the race of each registered voter. First, I conducted an ecological

regression analysis that relates rates of possession of acceptable SB 14 ID among

registered voters to the racial composition of Census areas. Second, I matched individual

voter records to data from a firm, Catalist, LLC, that uses a statistical model to predict the

race of individual registered voters.

6. Separate from the determination of who has acceptable SB 14 photo ID, I have also

been asked to examine historical rates of registration and voting among racial groups. I

did so in three ways: (i) using data from the Bureau of Census's Current Population

Survey (CPS); (ii) by ecological regression analysis relating registration and voting rates

---

[1] There are approximately 801,000 TEAM records (see Table V.3.A) that do not match to
any valid record in any applicable identification database. While none of these 801,000
TEAM records matched to any valid ID record on any applicable database, a relatively
small number matched to DPS records that are both invalid for SB 14 purposes (because
they are expired for more than 60 days or marked as surrendered, and as such, not in the
person's possession) and marked as deceased. Subtracting such records as deceased
reduces the number to 786,727.

to the racial composition of Census areas; and (iii) by analysis of TEAM voter history

data combined with the Catalist classification of the race of voters.

7.   In brief, I conclude that:

(a) There are 786,727 voters in Texas who do not possess acceptable SB 14 photo

identification, representing 5.8 percent of registered voters. That is, these are records on

TEAM that do not match to an identification databases and for which there is no

indication in the DPS data that the individual is deceased. Moreover, approximately

701,696 voters in Texas neither possess acceptable SB 14 photo ID nor qualify under SB

14 to apply for a disability-based exemption from showing ID at the polls.

(b) There is a racial disparity in the rates of possession of acceptable SB 14 photo ID

such that Black[2] registered voters are approximately twice as likely as Anglo registered

voters to lack acceptable photo ID and that Hispanic registered voters are approximately

fifty percent more likely than Anglo registered voters to lack acceptable SB 14 photo ID.

(c) In recent elections examined, rates of both voter registration and voter turnout are

lower for Blacks and Hispanics in Texas than Anglos.

8. I have been asked to respond to the reports of Professors Hood and Milyo as they

concern the state of academic research on voter identification laws and my analysis of the

TEAM data.

---

[2] Throughout, Black refers to individuals who are Black non-Hispanic.

II. Background and Qualifications

9. I am a professor of Government in the Department of Government at Harvard University in Cambridge, MA. Formerly, I was an Assistant Professor at the University of California, Los Angeles, and I was Professor of Political Science at the Massachusetts Institute of Technology, where I held the Elting R. Morison Chair and served as Associate Head of the Department of Political Science. I directed the Caltech/MIT Voting Technology Project from its inception in 2000 through 2004, am the Principal Investigator of the Cooperative Congressional Election Study, a survey research consortium of over 250 faculty and student researchers at more than 50 universities, and serve on the Board of Overseers of the American National Election Study. I am a consultant to CBS News' Election Night Decision Desk. I am a member of the American Academy of Arts and Sciences (inducted in 2007).

10. I have worked as a consultant to the Brennan Center in the case of *McConnell v. FEC*, 540 U.S. 93 (2003). I have testified before the U.S. Senate Committee on Rules, the U.S. Senate Committee on Commerce, the U.S. House Committee on Science, Space, and Technology, the U.S. House Committee on House Administration, and the Congressional Black Caucus on matters of election administration in the United States. I filed an amicus brief with Professors Nathaniel Persily and Charles Stewart on behalf of neither party to the U.S. Supreme Court in the case of *Northwest Austin Municipal Utility District Number One v. Holder*, 557 U.S. 193 (2009). I am consultant for the Rodriguez

plaintiffs in *Perez v. Perry*, currently before the U. S. District Court in the Western

District of Texas (No. 5:11-cv-00360), and the Gonzales intervenors in *State of Texas v.*

United States before the U.S. District Court in the District of Columbia (No. 1:11-cv-

01303); I consulted for the Department of Justice in *State of Texas v. Holder*, before the

U.S. District Court in the District of Columbia (No. 1:12-cv-00128); I consulted for the

Guy plaintiffs in *Guy v. Miller* in U.S. District Court for Nevada (No. 11-OC-00042-1B);

I consulted for the Florida Democratic Party in *In re Senate Joint Resolution of*

Legislative Apportionment in the Florida Supreme Court (Nos. 2012-CA-412, 2012-CA-

490); I am consultant for the Romo plaintiffs in *Romo v. Detzner* in the Circuit Court of

the Second Judicial Circuit in Florida (No. 2012 CA 412); I am consultant for the San

Antonio Water District intervenor in *LULAC v. Edwards Aquifer Authority* in the U.S.

District Court for the Western District of Texas, San Antonio Division (No. 5:12cv620-

OLG,); I am consultant for the Harris plaintiffs in *Harris v. McCrory* in the U. S. District

Court for the Middle District of North Carolina (No. 1:2013cv00949).

11. My areas of expertise include American government, with particular expertise in

electoral politics, representation, and public opinion, as well as statistical methods in

social sciences. I have authored numerous scholarly works on voting behavior and

elections, the application of statistical methods in social sciences, legislative politics and

representation, and distributive politics. This scholarship includes articles in such

academic journals as the Journal of the Royal Statistical Society, the American Political

Science Review, the American Economic Review, the American Journal of Political

Science, Legislative Studies Quarterly, the Quarterly Journal of Political Science,

Electoral Studies, and Political Analysis.   I have published articles on issues of election

law in the Harvard Law Review, Texas Law Review, Columbia Law Review, New York

University Annual Survey of Law, and the Election Law Journal, for which I am a

member of the editorial board. I have coauthored three scholarly books on electoral

politics in the United States, The End of Inequality: Baker v. Carr and the

Transformation of American Politics, Going Negative: How Political Advertising

Shrinks and Polarizes the Electorate, and The Media Game: American Politics in the

Media Age. I am coauthor with Ted Lowi, Ben Ginsberg, and Ken Shepsle of American

Government: Power and Purpose. My curriculum vita with publications list is attached to

this report.

12. As the Principal Investigator of the Cooperative Congressional Election Study and

the Harvard Election Data Archive, I have extensive experience with database

management, record linkage and database matching, data validation, and integration of

Census and electoral data. I have published articles in refereed journals on matching

survey data to voter files, on validation of voting records, and on statistical techniques for

analyzing aggregate election and population data.

13. I have been hired by the Department of Justice in this case.   I am retained for a rate

of $400 per hour, which is my standard consulting rate.

III.   Summary of Analysis and Findings

14. I have been asked to determine the number of Texas registered voters who lack

acceptable SB 14 photo ID and to determine whether there are disparities between the

percentages of Anglo registered voters and of Black or Hispanic registered voters who

possess such identification and those who neither possess such identification nor qualify

for SB 14's disability exemption from providing identification at the polls.

15. In order to do this, I conducted database matching and record linkage of the Texas

Election Administration Management system (TEAM) database to relevant State of

Texas and Federal databases.   SB 14 specifies 7 categories of state and Federal

identification that may be used when voting.[3] In addition, SB 14 allows persons with

certain Federally-documented disabilities to apply for an exemption from showing

identification at the polls.[4] The TEAM database records which voters have applied for

and received this disability exemption. Using standard methods for linking databases, I

matched records in the TEAM voter registration database to records of holders of Texas-

---

[3] Specifically, SB 14 requires that all in-person voters present one of the following in
order to cast a regular ballot: (1) a Texas Driver License (DL); (2) a Texas Personal
Identification Card (PID); (3) a Texas concealed handgun license (CHL); (4) a Texas
Election Identification Certificate (EIC); (5) a U.S. military identification card with
photo; (6) a U.S. citizenship certification (defined to include certificates of naturalization)
with photo; or (7) a U.S. passport. SB 14 requires that IDs with an expiration date be
current or expired for less than 60 days.

[4] To apply for the disability exemption, voters must submit written documentation stating
that they do not have an acceptable SB 14 photo ID and showing either that they have
been determined to be disabled by the Social Security Administration, or that they have a
disability rating from the U.S. Department of Veteran Affairs of at least 50%. As of
January 15, 2014, only 18 Texas voters had been approved for a disability exemption
from showing acceptable SB 14 ID.

issued forms of SB 14 ID contained in the Texas Department of Public Safety (DPS)

databases. Also, I developed protocols for matching the TEAM voter registration

database to databases of the Federal agencies that issue allowable forms of SB 14

identification, as well as the databases of the Federal agencies that make relevant

disability determinations.[5]

16. The matching algorithm produces a NO MATCH list, which consists of all records

on TEAM: a) for which no valid matching record could be found in any identification

database; b) which are not reported as deceased on any matching record in a DPS

database (including invalid for SB 14-purposes expired DPS records); and c) which are

not recorded in TEAM as having received a disability exemption.[6]   Each record on this

list is treated as an individual registered voter who lacks acceptable SB 14 photo ID.

The algorithm also produces a MATCH list, which consists of all records on TEAM for

which a valid matching record could be found on an identification database

17. The matching algorithm also produces a NO MATCH/NOT EXEMPTION

ELIGIBLE list, which consists of all records in TEAM: a) for which no valid matching

record could be found in any state or Federal identification database; b) which are not

reported as deceased on any matching record in a DPS database (including invalid for

---

[5] The Federal agencies that issue allowable SB 14 photo ID and which matched their
databases to TEAM are: the Department of Defense (DOD), Department of State (DOS),
Department of Veterans Affairs (VA), and the United States Citizenship and Immigration
Services (USCIS). The Federal agencies that make the relevant disability determinations
are the Social Security Administration (SSA) and the VA.
[6] Only 18 registered voters in TEAM were marked as having already received the
disability exemption.

SB 14-purposes expired DPS records); c) which are not recorded in TEAM as having received a disability exemption; and d) for which no matching record in the relevant Federal disability databases could be found. Each record on this list is treated as an individual registered voter who lacks acceptable SB 14 photo ID and does not qualify for SB 14's disability exemption.

18. Both the NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE lists are highly relevant to the question of the burden imposed by SB 14 on some Texas voters. Voters who qualified to apply for the disability exemption will not be able to avail themselves of that exemption on Election Day, unless they have taken steps to apply for and prove they qualify for the exemption.

19.   The TEAM data extracted on January 15, 2014, contains 13,564,416 records. Of the 13,564,416 records, 801,029 were not matched to any record in a State of Texas or Federal identification database or had not already received the disability exemption. Accounting for deceased individuals as indicated by DPS, the total number of records on TEAM is 13,487,594, and the NO MATCH list contains 786,727 voter records. (See Tables V.3.A and V.3.B)

20.   Of the 13,564,416 TEAM records extracted on January 15, 2014, 715,804 were not matched to any record in a State of Texas or Federal identification or disability database or had not already received the disability exemption.[7]   Accounting for deceased

9

individuals as indicated by DPS, the NO MATCH/DISABILITY EXEMPTION

ELIGIBLE list contains701,696 voter records. (See Tables V.4.A and V.4.B)

21. I performed two analyses of racial disparities in the incidence of inclusion on the NO

MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE lists.   First, the likelihood

that a registered voter is on the NO MATCH list or the NO MATCH/NOT EXEMPTION

ELIGIBLE is correlated with racial data from the Census Bureau, using ecological

regression. Ecological regression is widely used to measure the voting patterns of racial

groups in Voting Rights Act cases. (*Thornburg v. Gingles*, 478 US 30 (1986).) Second,

the likelihood that an individual record in TEAM is on either the NO MATCH list or the

NO MATCH/NOT EXEMPTION ELIGIBLE list is related to the individual's race, using

the classification of race of registered voters provided by Catalist, LLC, a data utility

company that provides election data.

22. These two forms of analysis for the NO MATCH list yield very similar results and

show that there are statistically significant[8] disparities between the rate at which Anglos

in TEAM are matched to identification databases and disability data records and the rates

at which Blacks and Hispanics are matched.   Analysis of the NO MATCH/NOT

EXEMPTION ELIGIBLE list shows that the racial disparities found as to the NO

---

[8] Throughout I report as "statistically significant differences" results for which a
statistical test yields a probability of less than 1 percent that the observed difference arose
by chance (i.e. that the true difference is 0). This corresponds to the observation that the
hypothesized value of 0 is outside of the 99 percent confidence interval for the observed
difference. In other words, the probability of observing a difference this large by chance
is less than 1 percent.

MATCH list are not alleviated upon considering eligibility to apply for the disability exemption.

23. Ecological regression based on the NO MATCH list estimates that 3.6 percent of registered Anglos do not match to a corresponding record in a state or federal identification database. By comparison, an estimated 11.9 percent of registered Blacks and 8.6 percent of registered Hispanics do not match. Stated differently, the rate of NO MATCH for black voters when measured through ecological regression is 231% higher than the rate of NO MATCH for Anglo voters. The rate of NO MATCH for Hispanic voters is 139% higher than for Anglo voters. (See Table VI.1.)

24. Analysis of the NO MATCH list using the Catalist racial classification estimates shows that 5.1 percent of registered Anglos do not match to a corresponding record in a state or federal identification database. By comparison, an estimated 9.2 percent of registered Blacks and 6.4 percent of registered Hispanics do not match.   Using the Catalist estimates, the rate of NO MATCH for black voters is 80% higher than the rate of NO MATCH for Anglo voters. The rate of NO MATCH for Hispanic voters is 25% higher than for Anglo voters.   (See Table VI.2.)

25. Analysis of the NO MATCH/NOT EXEMPTION ELIGIBLE list using ecological regression and Census racial data shows that 3.3 percent of registered Anglos do not match to a corresponding record in a state or federal identification database, nor in a Federal disability database. By comparison, an estimated 9.8 percent of registered Blacks

11

and 7.6 percent of registered Hispanics do not match in the relevant databases. (See Table VI.3.A, first column.) Analysis of the NO MATCH/NOT EXEMPTION ELIGIBLE list using the Catalist racial classification estimates indicates that 4.6 percent of registered Anglos do not match to a corresponding record in a state or federal identification database, nor in a Federal disability database. By comparison, an estimated 7.9 percent of registered Blacks and 5.8 percent of registered Hispanics do not match in the relevant databases. (See Table VI.3.B.)

26. Further analyses validate these findings and test the robustness of the results to alternative specifications of the pool of registered voters and the classification of individuals' races. The pattern of results remains consistent across these alternate analyses, with each analysis showing statistically significant racial disparities in rates of matching of records in TEAM to records in acceptable SB 14 photo ID databases between Anglo registered voters and Black and Hispanic registered voters.   (See Section VII.)

27. I also examine voting and registration patterns of racial groups in the State of Texas from 2006 to 2012 to determine whether racial differences in participation exist currently or historically.   I examine three data sets: TEAM data merged with Catalist racial classifications, the Current Population Survey (CPS) reports on voting and registration rates, and aggregate data on registration and voting and racial composition of populations.   All three data sets show that Anglos in the State of Texas register and vote at higher rates than minorities. (See Section VIII.)

IV. Data and Sources

A. Terminology

28.   TEAM Database. The official list of registered voters in the State of Texas, maintained by the Texas Secretary of State.   The Texas Election Administration Management (TEAM) System includes registered voter information such as name, address, date of birth, gender, and past elections in which the person voted.

Identification Database. A state or Federal list of individuals with a given form of identification, such as a Driver License or Passport, which includes names, addresses, and other information about the individuals.

Disability Database. A Federal list of individuals with a given disability status, which includes names, addresses, and other information about the individuals.

Record. A row in a database containing the information in that database for a specific person. Also called a case.

Field.   A column in a database corresponding to information about each of the records in the database, such as first names or dates of birth.

Identifier. A field or constructed combination of fields for a particular record in a database that can be used to identify another record likely to be the same person in a different database.

Unique Identifier.   An identifier that is unique to a given individual.

Record linkage. A process whereby a given record in one database is matched to one or more records in other databases using identifiers for individual records. Also called matching.

Match.   A record in a database found to have at least one matching identifier in a separate database.

No Match. A record for which no match is found in other databases.

Sweep. A search conducted of all records in one database using a specified identifier for matching records in another database.

MATCH List. Records for all currently registered voters in TEAM for which at least one valid matching record is found in an identification database on at least one sweep, and which are not matched to any deceased DPS record.

NO MATCH List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption and for which no valid matching record is found in any identification database on any sweep, and which are not matched to any deceased DPS record.

NO MATCH/NOT EXEMPTION ELIGIBLE List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption and for which no valid matching record is found in any identification or disability database on any sweep, and which are not matched to any deceased DPS record.

NO MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption, for which no valid

14

matching record is found in any identification or disability database on any sweep, and

which do not establish that the voter is qualified to vote by mail on account of age, and

which are not matched to any deceased DPS record.

B. Data Used in the Matching Process

State of Texas Databases

29. Counsel for the Department of Justice provided me with voter registration and

identification databases from the State of Texas.

30. Texas Voter Registration Data: Texas voter registration records were extracted from

the TEAM system on January 15, 2014.   That date serves as the date of the election for

purposes of this analysis. Any form of acceptable SB 14 photo ID that is unexpired on

that date or expired no earlier than sixty days prior to that date is deemed valid under SB

14 for purposes of this analysis.

31. State of Texas Photo ID Data: Records were extracted from the Texas Department

of Public Safety (DPS) databases for Driver Licenses (DL), Personal State Identification

Cards (PID), Licenses to Carry Concealed Handguns (CHL), and Election Identification

Certificates (EIC). The DPS DL, PID, CHL, and EIC databases also include a field that

indicates whether an individual may be deceased. The algorithm matches TEAM records

to identification records for each specific form of ID separately.   The DPS data

addressed in this report includes records that were provided in February 2014, as well as

records which I received in late July, 2014, but which I should have been provided in

February. It is my understanding that these records were not initially extracted because

of a coding error.

Federal Databases

32. I created a version of the TEAM database with identifiers used in matching already

formatted to facilitate matching to Federal databases. I developed a step-by-step

explanation of the algorithm, including data preparation steps, and developed model

computer code in STATA and SQL to guide the matches of the Federal databases.

Counsel for the United States conveyed these files to staff at the United States

Citizenship and Immigration Service, the United States Department of Defense, the

United States Department of State, the United States Department of Veterans Affairs, and

the Social Security Administration.   I had no direct access to Federal databases.[9]

C. Databases Used in Data Analyses

33. I analyze individual and aggregate-level data to estimate the relationship between

race and probability of possession of ID accepted under SB 14.   Individual-level

---

[9] Staff at each agency involved in the matching process completed a declaration
documenting the steps that they took in completing the matches. Those declarations are
provided as exhibits in the Appendix.

analyses rely on TEAM data matched to identification and disability databases and

Catalist racial classifications.

34. Aggregate-level analyses in Section VI sum the number of Matched and Not

Matched voters in each Census Block Group (BG).   I calculate the percent of registered

voters in each BG that are on the NO MATCH list.   Block Group is the lowest level of

aggregation for which Citizen Voting Age Population (CVAP) is released by the Census.

I correlated the percent NO MATCH with the percent of each racial group that are adult

citizens (CVAP) or are adults (VAP) at the BG level. The correlation allows me to

estimate the share of the CVAP or VAP for each racial group that lacks acceptable SB 14

photo ID. The correlations and ecological regressions reported in Section VI are of the

NO MATCH and MATCH data aggregated to the BG level for the CVAP.[10]   (See Tables

VI.1, VI.3.A, VI.4.A.)   Similar results hold for VAP, but CVAP is a closer

approximation to the eligible electorate. I also tested the robustness of the analysis at the

Census Tract level (a higher level of aggregation than BG), and found no substantive or

statistical difference in the results. I report results at the BG level.

35. Aggregate analyses in Section VII on rates of turnout and registration are at the

Voting Tabulation District (VTD) level as that is the level at which registration and

---

[10] In the case of VAP, I used the Census figures for the VAP for each group at the BG
level. In the case of CVAP, I use ACS data aggregated to the BG.

voting statistics are reported.   These analyses use Voting Age Population (VAP) from

the Census Enumeration and Citizen Voting Age Population (CVAP) from ACS.[11]

Voting and Registration Data

36. Analyses of historical voting and registration patterns use aggregate data from the

Texas Legislative Council, a nonpartisan State legislative agency. Voting and

registration data at the VTD level come from the website of the Texas Legislative

Council.[12]   Individual data on voter participation history in particular elections is also

used from the TEAM database.

Census Data

37. The Population and Voting Age Population in Voting Tabulation Districts (VTDs)

are collected by the 2010 Census Enumeration and come from the website of the Texas

Legislative Council.13

38. Aggregate data on the Citizen Voting Age Population (CVAP), at the Census Block

Group and VTD level, come from the five-year average of the American Community

Survey (ACS), 2008-2012.[14]

---

[11] For Block Groups that contain multiple VTDs, I apportion the CVAP of a racial group
in a Block Group based on the percent of the VAP of that group in a Block Group that
resides in each VTD.

[12] ftp://ftpgis1.tlc.state.tx.us/elections/

[13] ftp://ftpgis1.tlc.state.tx.us/2011_Redistricting_Data/VTDs/

18

39. I analyze Census data on the racial composition of the electorate and the rate with which registered voters are deemed to possess acceptable SB 14 identification. I perform Ecological Regression at the Block Group (BG) level to measure the rate of NO MATCH of each racial group.

40. The Census Bureau conducts the Voting and Registration supplemental to the CPS each November following federal elections. I examine the 2006, 2008, 2010, and 2012 CPS.[15]

Catalist Database

41. The United States contracted with Catalist, LLC, to obtain additional data on voter registration records in the State of Texas. I use Catalist data (i) to test the robustness and validity of findings, (ii) to obtain estimated classifications of the race of individuals (estimates that are based in part on local area Census demographics and frequencies of names), (iii) to examine voting history of individuals, and (iv) to obtain geocoding of each registration record and Catalist deadwood, deceased, and NCOA information.

42.   Catalist maintains data on voter registrations and vote history. Catalist retains the TEAM VUID, which permits linkage between the Catalist data and the TEAM data.

---

[14] *See* http://www.census.gov/acs/www/data_documentation/2012_release/
[15] http://www.census.gov/hhes/www/socdemo/voting/publications/p20/index.html

43. Catalist augments official voter registration data with indicators regarding whether a

voter has moved as reported through the National Change of Address (NCOA) data from

the United States Postal Service, information on whether a voter is deceased using data

from the Social Security Administration and private vendors, and information identifying

potentially obsolete records (or deadwood) based on factors including participation in

past elections. To analyze the validity and robustness of findings, I use the Catalist flags

for NCOA, deceased, and deadwood records to construct subsets of the pool of

registration records.

44. Catalist provides a classification of the race of each individual in the TEAM

database and a score for the confidence in that classification. Catalist's race

classification is based on the frequency of specific last names in the population and the

frequency of racial groups in local areas (Census block groups), and is a refinement on

the area-based estimates underpinning the ecological regression analysis frequently used

in Voting Rights Act cases.

45. Catalist data are widely used in academic research on registration and voting and

have been vetted for publication in peer-reviewed journals. Catalist data on

demographic characteristics of the electorate, including age, gender, and race, have been

vetted and published in peer-reviewed journals.[16] Academic researchers use the Catalist

---

[16] See for example, Ansolabehere, Stephen, and Eitan Hersh. "Validation: What big data
reveal about survey misreporting and the real electorate." *Political Analysis* (2012):
mps023. Ansolabehere, Stephen, Eitan Hersh, Kenneth Shepsle. "Movers, Stayers, and
Registration: Why Age is Correlated with Registration in the US." *Quarterly Journal of
Political Science* 7, no. 4 (2012): 333-363. Ansolabehere, Stephen, and Eitan Hersh.

database to identify the population of registered voters in the US and to conduct random

sample surveys of the population of registered voters and experimental research on voter

participation.[17]

46. I verify the validity of the inferences for the full Catalist analysis by examining the

subset of records in Catalist with very high confidence in the racial classification.

Catalist has very high confidence in areas where there is a fairly homogeneous population

and for individuals with sufficiently distinctive names.

V. Matching Process

A. Description of the Matching Process

47.   The matching process conducts record linkage for individual records in TEAM to

any record in each database corresponding to a form of identification accepted under

---

"Gender, Race, Age and Voting: A Research Note." *Politics and Governance* 1, no. 2 (2013): 132-137. Garcia-Castañon, Marcela, Alison D. Rank, and Matt A. Barreto. "Plugged in or tuned out? Youth, race, and Internet usage in the 2008 election." *Journal of Political Marketing* 10, no. 1-2 (2011): 115-138.

[17] See, for example, Nickerson, David W., and Todd Rogers. "Political Campaigns and Big Data." *The Journal of Economic Perspectives* 28, no. 2 (2014): 51-73. Dale, Allison, and Aaron Strauss. "Don't forget to vote: Text message reminders as a mobilization tool." *American Journal of Political Science* 53, no. 4 (2009): 787-804. Bennion, Elizabeth A., and David W. Nickerson. "The Cost of Convenience An Experiment Showing E-Mail Outreach Decreases Voter Registration." *Political Research Quarterly* 64, no. 4 (2011): 858-869. Ansolabehere, Stephen, and Eitan Hersh. "Validation: What big data reveal about survey misreporting and the real electorate." *Political Analysis* (2012).

SB 14 or in a database reflecting those eligible to apply for the disability exemption. The matching algorithm proceeds in four parts.

Database Preparation.   Databases are prepared and standardized.

Creation of Identifiers.   Identifier values used to link records in one database to records in another database are constructed by combining multiple individual fields.

Record Linkage and Matching. One-to-many matches are conducted between the databases. That is, the algorithm matches each unique identifier on the TEAM database to all records on the identification database that have the corresponding value of the identifier.

Data Gathering. Appended to the TEAM data are fields indicating every match found of a record on the TEAM database to a record on a state or Federal identification database or Federal disability databases.

48. Each of the four parts is divided into multiple stages, which are in turn divided into concrete steps.   Detailed procedures for the implementation of the algorithm were prepared and presented to all parties in the litigation and to the Federal agencies. These memorandums provide detailed, step-by-step documentation of the matching process. Rather than reproduce the detailed steps here, the memorandums presented to all parties and the Federal agencies are appended to this report in the appendix. This section describes the stages of the Matching Process at a general level.

49. The result of this methodology is to produce a MATCH list, a NO MATCH list, and

a NO MATCH/NOT EXEMPTION ELIGIBLE list, as described above.

B. Features of the Matching Algorithm

50. The first phase of the matching algorithm, Database Preparation, standardizes the

coding of database fields to facilitate matching.   Different databases store the fields in

different ways.   For example, Gender is coded 1 or 0 in some databases and M or F in

others. The database preparation in the algorithm standardizes the coding of names by

removing spaces, hyphens, and other characters; standardizes dates of birth and gender

codes, and identifies invalid or missing values (such as 111111111 for Social Security

Numbers). Prior research has shown that standardization of fields, removal of duplicate

cases, and definition of missing or invalid values in each field greatly improves the

quality of matching.[18]

51. The second part of the algorithm develops multiple identifiers for purposes of record

linkage. This general approach is widely used in the field of record linkage and database

matching, especially in health research and marketing, and has been determined by past

---

[18] William E. Winkler, "Methods for Evaluating and Creating Data Quality," Information
Systems 29 (October, 2004), 531-550.
http://www.sciencedirect.com/science/article/pii/S030643790400002X
Max G. Arellano and Gerald I. Weber, "Issues in Identification and Linkage of Patient
Records Across an Integrated Delivery System," Journal of Health Care Information 12,
no. 3 (1998): 43-52
http://sce.umkc.edu/~leeyu/Mahi/medical-data5.pdf

studies to yield a very high rate of correct matches.[19] The algorithm builds identifiers by combining fields related to Address, Date of Birth, Gender, Name, Social Security Number, and Texas Driver License Number.

52. In total 13 different identifiers were constructed in the TEAM database and in the corresponding State and Federal databases.[20] Each identifier corresponds to a particular combination of fields. For example, Combination A consists of First Name, Last Name, Date of Birth, Gender, Street Number, and 5-digit ZIP Code.   A sample version of Combination A for a man named John Smith, born on January 1, 1960, and living at 100 Main Street in the ZIP Code 78610 would be JOHNSMITH01011960110078600.   A chart of the identifiers created for matching is provided below in Table V.1. Appendix Tables A.V.1-A.V.2 provide statistics on the completeness and uniqueness of combinations of fields.[21]

---

[19] See Simon, Michael S., Beth A. Mueller, Dennis Deapen, and Glenn Copeland. "A comparison of record linkage yield for health research using different variable sets." *Breast cancer research and treatment* 89, no. 2 (2005): 107-110. Sweeney, Latanya, "Computational Disclosure Control for Medical Microdata: The Datafly System, Record Linkage Techniques 1997, Chaper 11. Pp. 442-453. Sweeney, Latanya, Matching Known Patients to Health Records in Washington State Data. Harvard University. Data Privacy Lab. 1089-1. June 2013.

[20] For the Department of State matches of passport holders, additional identifiers were developed to address specific features of the way that the Passport database stores name information.   Those additional identifiers are discussed in a declaration from DOS, which is attached in the appendix.

[21] These tables were not revised after receiving the 3.1 million supplemental DPS records. They correspond to the data extracted on January 15, 2014.

Table V.1. Combinations of Fields Used as Matching Identifiers	
Combination Code	PRIMARY MATCHES
A	First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number
B	Last Name + Gender + DOB + Residential ZIP + Residential Street Number
C	Gender + DOB + Residential ZIP + Residential Street Number
D	First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number
E	First Name + Last Name + Gender + Residential ZIP + Residential Street Number
F	First Name + Last Name + Gender + DOB
M	Texas Driver License Number (where available)
	SECONDARY MATCHES
G	First Name + Middle Initial + Last Name + DOB
H	Last 4-Digit SSN + DOB + Residential ZIP
I	Last 4-Digit SSN + First Name + Last Name + DOB
K	First Name + Last Name 1 + Middle Initial + DOB
L	First Name + Last Name 2 + Middle Initial + DOB
SSN	9-Digit Social Security Number

25

53. The third stage of the process, the Record Linkage and Matching phase, conducts

one-to-many matches[22] and performs multiple sweeps to guard against false negatives

(non-matches that should be matched).   An example of false negatives that this approach

guards against is a typographical error in the spelling of a person's name as recorded on

TEAM but not in the DPS driver license database.   Such a typo would create an

inconsistency between TEAM and DPS if matches were conducted only on identifiers

that included name. In addition to searching on identifiers that contain name fields the

algorithm searches on identifiers constructed from combinations that do not include name

elements, such as Date of Birth, Gender, Address, and Social Security Number. The

algorithm will match the record on the identifiers that do not contain each of these

categories of fields, thus avoiding non-matches due to typographical errors, nicknames,

missing fields, and other inconsistencies between databases. A record is determined to

have found a match if a given identifier in TEAM is identical to at least one

corresponding identifier in an identification database.   The frequencies of matches of

individual records in TEAM to specific identifiers in each of the state and federal

---

[22] As used here, a "one-to-many" match means that for each TEAM record, a match will
only be attempted for a particular identifier when that identifier is unique to a single
TEAM record. However, that unique identifier will be deemed to have matched to any
identical identifier in an identification or disability database, even if the identifier is not
unique on the identification or disability database. As an example, consider a simplified
identifier not used here: First Name + Last Name + Date of Birth. If there are two people
named JOHN SMITH with a birth date of January 1, 1960 in TEAM, no match is
attempted on First Name + Last Name + Date of Birth because the TEAM identifier is
not unique. This guards against false positives. On the other hand, if there is only one
JOHN SMITH in TEAM born on January 1, 1960, and there are two JOHN SMITHs
born on January 1, 1960 in the DPS driver license database, the match will be attempted,
and the TEAM record is considered to have matched. All that matters is that the
registered voter John Smith is deemed to hold a driver license—there is no need to
distinguish between the two DPS records.

identification and disability databases are presented in the appendix.   See Appendix

Tables A.V.3 to A.V.5.

54. As shown in the chart below, the algorithm conducts two sorts of sweeps through the

data to find matching records. The Primary Sweeps match on Combinations A-F and M,

and are run on all TEAM records.   The Secondary Sweeps are conducted on

Combinations G – L for the TEAM records not matched in the Primary Sweeps.   For

Federal databases, the Primary Sweeps are run against all qualifying Federal records with

Texas addresses, while the Secondary Sweeps are run both against Texas-only records, as

well as against the nationwide universe of the relevant Federal dataset.

		Matching Combinations
Texas DPS Databases	Primary Sweeps (All TEAM records)	Combination A: First name + Last name + Gender + DOB + Street number + ZIP Combination B: Last name + Gender + DOB + Street number + ZIP Combination C: Gender + DOB+ Street number + ZIP Combination D: First name + Last name + Street number + ZIP Combination E: First name + Last name + Gender + Street number + ZIP Combination F: First name + Last name + DOB + Gender Combination M: Texas Driver License Number
	Secondary Sweeps (TEAM records with no primary match)	Combination G: First name + Last name + Middle Initial + DOB Combination H: DOB + ZIP + SSN4 Combination I: First name + Last name + DOB + SSN4 Combination K: First name + Last name 1 + Middle Initial + DOB Combination L: First name + Last name 2 + Middle Initial + DOB23 SSN: 9-digit Social Security Number
Federal Identification and Disability Databases	Primary Sweeps (All TEAM records against Federal records with a Texas address)	Same as primary sweeps for DPS databases, except for Texas Driver License Number (Combinations A-F)
	Secondary Sweeps (TEAM records with no primary match against Federal records with a Texas address)	Same as secondary sweeps for DPS databases (Combinations G-L and SSN)
	Nationwide Sweeps (TEAM records with no primary or secondary match against nationwide Federal records)	All sweeps without address criteria (Combinations F, G, I, K, L, and SSN)

---

[23] "Last name 1" is the first half of a hyphenated last name, and "Last name 2" is the second half of a hyphenated last name. Combinations K and L in TEAM are each matched against Combination G, Combination K, and Combination L in the identification and disability databases for a total of six matching sweeps.

28

55. In the last phase of the matching process, the Data Gathering phase, the results of all

matching sweeps are recorded for each individual TEAM record. This stage also appends

indicators of deceased records from the Texas DPS data to TEAM.   In the analysis

performed after receiving the supplemental 3.1 million records, I proceeded as follows.

For the complete DPS list of drivers license holders and state ID holders (including those

both from the initial production and the supplemental production), I subset the complete

list to anyone with a deceased flag in the database. This subset consists of all records

with deceased flags, including records that are expired or that have card statuses

indicating that the license has been surrendered and, thus, cannot be used for voting. I

then performed the same matching algorithm between this list and TEAM as I did for IDs

in DPS.   Similarly I used the same process to match deceased records from the CHL list

to TEAM. Any TEAM record that matched on any indicator to a deceased flag in DPS or

CHL was treated as deceased.   These records are not treated as registered voters in

subsequent analysis. In my initial report, I only treated as deceased TEAM records that

matched to a valid DPS record that was marked as deceased. This methodology could

only remove deceased voters from the MATCH List. The additional step undertaken

here—examining the entire universe of DPS records (both valid and invalid for SB 14

purposes) to locate deceased records—uses all possible available information from DPS

to identify likely deceased records on TEAM so that they can be removed from either the

MATCH List or the NO MATCH List.

29

56. The algorithm developed for DOJ in the present case goes beyond the matching

algorithm developed for the expedited Section 5 proceedings in State of Texas v. Holder

in which I also testified on behalf of the United States.[24]   First, the TEAM database is

now matched to all relevant state and federal databases.   In the section 5 proceeding,

time constraints prevented an assessment of federal identification and disability data.

Second, the algorithm ensures that persons that DPS considers to be dead are not

included on the NO MATCH list.   Instead, deceased voters are identified after matching

has occurred.[25]   Third, by using multiple identifiers, the algorithm is developed to be

sensitive to variations in names, such as nicknames and compound names, to

typographical errors, and to missing information.   Fourth, by matching on identifiers

constructed from a larger number of categories of fields (three or four rather than two),

the algorithm is more precise in determining which records link and more exhaustive in

the search for linkages.

57. The data analysis presented in the Validation section of this report examines the

robustness of the results to alternative racial classifications.

C. MATCH, NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE Lists

Match Rates

---

[24] The Court in *Texas v. Holder* chose not to rely on any of the expert testimony
presented. 888 F. Supp. 2d 113, 134-138 (D.D.C. 2012).
[25] Where a TEAM record has matched to a DPS record that the State of Texas has
marked as deceased, recording this information allows matched records for dead people
to be excluded from further analysis of the population of voters actually affected by SB
14's requirements.

58. The implementation of the algorithm developed for the United States in this case matched the entire TEAM database to 10 different state and federal databases.   Table V.2 below lists the number of records in TEAM that matched to each state or Federal database using that algorithm, as well as the percent of TEAM records overall and in each racial group that match to each identification or disability database. (Note: This table corresponds to all matched TEAM records, before removal of TEAM records indicated as deceased by DPS, and shows the total rate of matching between TEAM and each form of identification and database.)

59. The most commonly held form of identification is a State of Texas Driver License, followed by a United States Passport.   Just over 83.8 percent of records in TEAM matched to the DPS Driver License list, while 42.3 percent of records in TEAM matched to the DOS passport database. The next most common form of ID is a DPS Personal (or State) ID, held by 9.2 percent of those in TEAM.

60. Table V.2 shows that the rates with which records match to the databases varies with race. The percentage of records on TEAM matched to Driver Licenses and Passports is much higher for Anglos than for Blacks and Hispanics. 86.8 percent of Anglos on TEAM match to a record on the DPS Driver License database, compared with 74.7 percent for Blacks and 80.3 percent for Hispanics.   45.5 percent Anglos on TEAM match to a record on the DOS passport database, compared with 24.7 percent for Blacks and 37.7 percent for Hispanics.

Table V.2. Number of Matches of TEAM Records to State and Federal Databases
Overall and By Racial Group, using Catalist Racial Estimates
(Percent of TEAM Records that Match to a Given ID or Disability Database)

Database	Race				
State of Texas ID Databases	White	Black	Hispanic	Other	All
Driver License	7,201,903 (86.8%)	1,283,720 (74.7%)	2,455,518 (80.3%)	423,139 (85.8%)	11,364,280 (83.8%)
Personal ID	417,122 (5.0%)	312,017 (18.1%)	495,983 (16.2%)	28,666 (5.8%)	1,253,788 (9.2%)
Concealed Handgun License	588,087 (7.1%)	57,129 (3.3%)	72,953 (2.4%)	14,839 (3.0%)	733,008 (5.4%)
EIC	69	43	51	0	163
Federal ID Databases					
DOS	3,776,207 (45.5%)	424,682 (24.7%)	1,151,608 (37.7%)	378,666 (76.8%)	5,731,163 (42.3%)
DOD	427,191 (5.2%)	81,688 (4.8%)	116,460 (3.8%)	13,015 (2.6%)	638,354 (4.7%)
USCIS	106,051 (1.3%)	45,005 (2.6%)	373,576 (12.2%)	210,454 (42.7%)	735,086 (5.4%)
VHA (VIC)	186,695 (2.3%)	49,179 (2.9%)	57,635 (1.9%)	2,496 (0.5%)	296,005 (2.2%)
Federal Disability Databases					
SSA: Disability	419,065 (5.1%)	167,980 (9.8%)	202,368 (6.6%)	14,925 (3.0%)	804,338 (5.9%)
VBA: Disability	118,883 (1.4%)	31,952 (1.9%)	35,743 (1.2%)	1,938 (0.4%)	188,516 (1.4%)

61. The rate at which records on TEAM match to DPS Personal ID or USCIS databases

is higher for minorities than for Anglos.   Of Anglos on TEAM, 5.0 percent match to

records on the DPS Personal ID database and 1.3 percent match to the USCIS database of

holders of certificates of citizenship and naturalization. Of Blacks on TEAM, 18.1

percent match to the Personal ID database and 2.6 percent match to the USCIS database.

Of Hispanics on TEAM, 16.2 percent match to the Personal ID database and 12.2 percent

match to the USCIS databases.   However, both the DPS Personal ID and USCIS

Certificates of Citizenship and Naturalization are much less commonly held than Driver

Licenses and Passports.

Sizes of the MATCH, NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE
Lists

62.   Of the 13,564,416 records in the TEAM database, 12,163,160 matched to at least

one record corresponding to acceptable SB 14 photo ID issued by the State of Texas, and

6,326,122 records matched to at least one record corresponding to acceptable SB 14

photo ID issued by the Federal government. (See Table V.3.A) Most of the records

matched to the Federal databases also matched to a State of Texas identification database.

No valid matching record was found on any of the state or Federal identification

databases and no disability exemption was granted for 801,029 records on the TEAM

database, approximately 5.9 percent of all records in TEAM. No matching record was

found for 715,804 records in the TEAM database matched to a state or Federal

identification database or to a Federal disability database. (See Table V.4.A) These

numbers include TEAM records that can be matched to DPS records flagged as deceased, as discussed in paragraph 53.

63. Accounting for matched voters that DPS data indicate as deceased from the number of records in the TEAM database reduces the universe of registered voters from 13,564,416 to 13,487,594.   I define this set, which removes from further analysis voters that the State of Texas data from DPS indicate to be deceased, as the Baseline Universe of Registered voters.   After removing TEAM records that matched to a deceased record from TEAM overall leaves 786,727 voters on the NO MATCH list, out of a the 13,487,594 records that I describe as the Baseline Universe on TEAM.   Again, this means that 5.8 of Texas registered voters for whom there is no indication in a state database of being deceased are on the NO MATCH list.   Likewise, there are 701,696 records on the NO MATCH, NOT EXEMPTION ELIGIBLE List, representing 5.2 percent of the Baseline Universe on TEAM.

VI.   Results: Analysis of Racial Disparities

64. This section analyzes how the rate of non-matched records between TEAM and valid SB 14 ID databases varies across racial groups.

A. Census Racial Data and Possession of ID

65. The American Community Survey (ACS) conducted by the Census Bureau provides estimates of the racial composition of the electorate. Using the 5-year average of the survey from 2008 to 2012, the ACS provides estimates of the CVAP of various racial groups at the block group level and tract level.

66. I aggregated the Match and No Match lists (which remove records indicated by DPS as deceased) to the block group level, the lowest geographic level at which Census reports ACS CVAP numbers.[26]   Within each block group, I computed counts of the numbers of registered voters and the number of registered voters who did not match any identification database and the number of registered voters who were matched to at least one identification database.   I then computed the percentage of registered voters in each block group who were not matched to any record.

67. I used ecological regression and homogeneous block group analyses[27] to estimate the percentage of Black citizens of voting age, Hispanic citizens of voting age, and Anglo citizens of voting age for whom a matching record to an identification database was found. Ecological regression estimates the relationship between Percent No Match and Percent of CVAP who are Anglo, Black, or Hispanic, enabling me to estimate the percent of each group who match to an ID database. Homogeneous block group analysis examines the subset of block-groups where all or almost all of the adult citizens are of one race—in this analysis, at least 80 percentage of a given race. Within such racially

---

[26] I conducted similar analyses at the Census tract level and discovered the same pattern of results.   Because Block Group is a lower level of aggregation I present that here.
[27] In other contexts data are at the precinct level, so homogeneous block group analysis is also called homogeneous precinct analysis.

homogeneous areas, I then compute the percent of registered persons for whom a

matching identification record was found or not found; this enables me to estimate the

percentage of that group who has (or does not have) an SB 14 ID. Both ecological

regression analysis and homogeneous block group analysis are well-established statistical

procedures relied upon in voting rights cases, where they are often used to measure

racially polarized voting and cohesiveness of voting of racial groups.

68. Table VI.1 presents the ecological regression and homogenous block group analysis

estimates of the rate with which No Match was found among each racial group. The

ecological regression estimates indicate that no match between TEAM and a state or

Federal identification database was found for approximately 4 percent of Anglos, 12

percent of Blacks, and 9 percent of Hispanics. The difference between the Black and

Anglo rate of non-matching, then, is 8 percentage points, and the difference between

Hispanics and Anglos is approximately 5 percentage points. In other words, Hispanics

are twice as likely as Anglos to be on the NO MATCH list, and Blacks are three times as

likely as Anglos to be a NO MATCH. (That ratio is reflected in the Relative Rate of NO

MATCH in Table VI.1 and subsequent tables.)

69. The homogeneous block group analyses in Table VI.1 are similar to the ecological

regression estimates. 4.5 percent of Anglo registered voters were estimated to have no

matching identification record on a state or Federal database. That figure was nearly 12.3

percentage points among Blacks and 8.8 points among Hispanics. In other words, the

difference in the rate of NO MATCH equaled 7.8 points between Blacks and Anglos and

36

4.3 points between Hispanics and Anglos. Again, in the homogeneous blocks, Hispanics are twice as likely as Anglos, and Blacks are almost three times as likely as Anglos to not match to a valid record in a qualifying identification database.

70. The results from the analysis of the homogeneous block groups and the ecological regressions are highly unlikely to have arisen by chance. The observed differences across the groups are statistically significant at the confidence levels generally used by social scientists.

B. Catalist Racial Data and Possession of ID

71. Analysis of individual level data using the Catalist classification of race yields similar results to the aggregate analyses presented in part VI.A.   The Baseline Universe of Registered Voters, which consists of all currently-registered voters in TEAM – after removing those who matched a record marked as deceased in a DPS ID file – has 13,487,594 records. Of these, 8,246,016 are classified as Anglo according to Catalist's estimates; 1,707,769 are Black; 3,042,497 are Hispanic; and 491,312 are Other Races. (See Table VI.2.)

72. The rate of non-matches between TEAM and identification databases varies by race. Of records identified as Anglo in the Baseline Universe of Registered Voters, 5.1 percent had no matching record in state or Federal identification databases. By comparison, no

37

matching records were found for 9.2 percent of people identified as Black and 6.4 percent of people identified as Hispanic. (See Table VI.2.)

73. The differences in rates of matching and non-matching across racial groups are statistically significant at the confidence levels normally used by social scientists. The difference between Blacks and Anglos in the rate of non-matching is 4.1 percentage points.   Such a difference is highly unlikely to have arisen by chance given the degree of precision of the estimates.[28] Likewise, the difference between Anglos and Hispanics in the rate of non-matching is 1.3 percentage points, which is highly unlikely to have arisen by chance.[29] Using the Catalist racial classification, Blacks are 80 percent more likely to be a NO MATCH and Hispanics are 25 percent more likely to be a NO MATCH.

C. Eligibility for Exemption and Vote-by-Mail

74. The analysis in Table VI.2 does not account for the fact that those with a Federally-determined disability may qualify and apply for an exemption.   Also, those over 65 may vote by mail without showing ID, but must still show ID if they vote at a polling place.

75. Table VI.3 parallels the analysis in Table VI.2, but classifies as MATCHES records for which the voter could apply for the disability exemption or vote by mail because of

---

[28] The 99 percent confidence interval for that difference in proportions is +/- 0.1 percentage points, so the differences are statistically different from 0.
[29] I use a 99 percent confidence level.   That is the probability of observing a difference this large by chance is less than 1 percent.

age.   Version VI.3.A presents the ecological regression results for this analysis using

Census' ACS data, and VI.3.B presents the results of the analysis using Catalist racial

data. The first column of each table shows the relationship between Race and percent on

the NO MATCH/NOT EXEMPTION ELIGIBLE.   That list consists of those not

matched to a state or Federal identification database or to a Federal disability database.

The second column of each table shows the relationship between Race and percent on the

NO MATCH/NOT AGE VOTE-BY-MAIL ELIGIBLE.   That list consists of those not

matched to a state or Federal identification database and under 65 years of age.   The

third column of each table shows the relationship between Race and percent on the NO

MATCH/ NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE.

That list covers people not potentially exempt because of disability, who are under 65,

and also not matched to a state or Federal ID. For this analysis the Baseline of

13,487,594 records is used as the pool of registered individuals.

76. The ecological regression analysis estimates that 2.4 percent of Anglos were NO

MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE.   By

comparison, 7.7 percent of Blacks and 5.9 percent of Hispanics were estimated to be NO

MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE.

(See the last column of Table VI.3.A.)   The differences between the rates of NO

MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE are

statistically significantly higher for minorities than for Anglos.   And these estimates are

very similar to those in Table VI.1.

77.   Among records classified as Anglo by Catalist's estimates, 3.3 percent were not matched to any record in any identification or exemption database and were under 65. By comparison, a NO MATCH and no potential exemption were found for 6.4 percent of people classified as Black and 4.6 percent of people classified as Hispanic. Similar patterns hold separately for those people who do not qualify for the disability exemption or for those under 65. See Table VI.3.B. These estimates are very similar to those in Table VI.2.

78.   The difference between Black and Anglo rates is 3.1 percentage points, and the difference between Hispanic and Anglo rates is 1.3 percentage points. Both differences are statistically significant at the confidence levels generally used by social scientists. Blacks are twice as likely as Anglos to have neither a MATCH nor qualify for an exemption; Hispanics are nearly forty percent more likely to have neither a MATCH to a state or federal identification record nor qualify for an exemption.

80. The aggregate and individual-level data are remarkably consistent. They show statistically significant differences between the rate with which Anglos and the rates with which Blacks and Hispanics on the TEAM database are matched to state and federal identification databases or are eligible for an exemption.   The difference between Blacks and Anglos is in the range of 3 to 8 percentage points and the difference between Hispanics and Anglos is in the range of 1 to 5 percentage points.

D.   Voting Rate among Registrations with No Match to an ID

81. The NO MATCH list contains 2.4 percent of all registered voters on TEAM who voted in 2012 and 1.7 percent of all registered voters who voted in 2010.

82. Table VI.4.A shows that the NO MATCH rate among voters is higher among minorities than among Anglos.   Ecological regression estimates that the NO MATCH rate among Anglos who voted in was 1.0 percent in 2010 and 1.7 percent in 2012.   The NO MATCH among Blacks who voted was 4.4 percent in 2010 and was 6.2 percent in 2012. The NO MATCH among Hispanics who voted was 2.7 percent in 2010 and was 3.1 percent in 2012.   The rates of NO MATCH among Black and Hispanic voters are statistically significantly higher than for Anglos.

83. Table VI.4.B provides a parallel analysis using the classification of individual race provided by Catalist.   Table VI.4.B shows that people classified using the Catalist data as Blacks who voted in 20102 were twice as likely to be NO MATCH or to be a NO MATCH AND NOT EXEMPTION ELIGIBLE as Anglos. The gross percentage point differences are in the range of 1 to 2 percentage points, depending on the election.   These differences are statistically significantly.   The racial classification provided by Catalist shows difference in rates of NO MATCH or NO MATCH AND NOT EXEMPTION ELIGIBLE between Hispanics who voted in 2010 and Anglos who voted in 2010 of one-half of one percentage point and no difference between these groups in 2012.   The difference in 2010 is statistically significantly different from 0.

41

VII. Validation

84. This section determines the sensitivity of results to alternative specifications of the pool of registered voters and to alternative classifications of race.

A. Variations in Universe of Registered Voters

85. The TEAM database is the official list of registered individuals in the State of Texas. Even so, there may be questions raised about the currency of some records in that database.   All states that have voter registration will have at least some registrations that are out-of-date or invalid but still on the rolls, for a wide variety of reasons.   I constructed alternative formulations of the list of registered voters to examine whether such concerns could affect the inferences drawn regarding racial disparities in rates of acceptable SB 14 ID possession.

86. This section examines variations in the pool of registered voters achieved by omitting records from TEAM for which there is some indication that the record may no longer be current or valid according to information from Catalist or internal to the TEAM or DPS databases.   Omitting these cases changes both the numerator and the denominator of calculations of the percent who are matched or not.

87. Table VII.1 presents the rates with which racial groups match to the relevant state

and federal databases under different constructions of the set of registered voters.   Table

VII.1.A presents ecological regression analyses using Census racial data, and Table

VII.1.B presents analyses using Catalist racial data.

 Catalist Deceased, Deadwood, and NCOA flags

88. The Catalist database includes indicators of whether an individual is deceased, is

deadwood (an obsolete record), or has a National Change of Address application on file

with the U.S. Postal Service, indicating that the individual has moved.[30] Each of these

categories provides evidence that a given registration record may no longer be current. I

performed two sorts of analyses using the Catalist Deadwood flags.   First, I performed

ecological regression analyses. I removed the records that Catalist flagged as deadwood

from the MATCH and NO MATCH lists, aggregated the data to the Block Group level,

and then performed ecological regression on the rate of NO MATCH (without those

indicated by Catalist as deadwood) on Census racial data. Second, I analyzed rates of

NO MATCH across racial groups using the classification of racial groups provided by

Catalist. Column 1 in Table VII.1.A and VII.1.B presents the overall NO MATCH figure

and the rate of NO MATCH among racial groups excluding records flagged by Catalist as

deceased, deadwood, or NCOA are excluded from TEAM.

---

[30] NCOA flags do not distinguish between in-county, in-state, and out-of-state moves.

89. Column 1 of Table VII.1.A shows that the NO MATCH rate among voters is higher among minorities than among Anglos after excluding records flagged by Catalist as deceased, deadwood, or NCOA.   Ecological regression estimates that the NO MATCH rate among Anglos in this subset of the data was 3.4 percent.   The NO MATCH among Blacks in this subset of the data was 11.3 percent. The NO MATCH among Hispanics in this subset of the data was 8.5 percent.   The rates of NO MATCH among Black and Hispanic voters are statistically significantly higher than for Anglos.

90. In Table VII.1.B there are significant racial differences in rates of NO MATCHES after removing from TEAM records that Catalist's data indicate as deceased, Deadwood, or NCOA and using the racial classification provided by Catalist.   Among records that the Catalist race estimates classified as Anglo, 4.9 percent were not matched to any record in any identification database. By comparison, no matches were found for 9.0 percent of people classified by Catalist as Black and 6.3 percent of people classified as Hispanic.

91. The difference between Black and Anglo rates of no matches is 4.1 percentage points, and the difference between Hispanic and Anglo rates of no matches is 1.4 percentage points. Both differences are highly unlikely to arise by chance.

Suspense Voters

92.   The TEAM database distinguishes Active and Suspense (or inactive) voters. A suspense voter is still legally registered but may be dropped from the registration list for reasons of non-voting or non-response to election office communications.[31] Column 2 in Tables VII.1.A and VII.1.B present the rates of NO MATCH among racial groups excluding Suspense Voters from the pool of Registered Voters.

93. Column 2 of Table VII.1.A presents the ecological regression estimates of non-matching rates for the racial groups after removing Suspense (or inactive) voters. Ecological regression estimates that the NO MATCH rate among Anglos in this subset of the data was 2.9 percent.   The NO MATCH among Blacks in this subset of the data was 10.2 percent. The NO MATCH among Hispanics in this subset of the data was 7.9 percent.   The rates of NO MATCH among Black and Hispanic voters are statistically significantly higher than for Anglos.

94.   Column 2 of Table VII.1.B presents the analysis of non-matching rates for the racial groups after removing Suspense (or inactive) voters using individual-level estimates of race provided by Catalist. Among records that Catalist estimates classified as Anglo, 4.3 percent were not matched to any record in any identification database. By comparison,

---

[31] The Suspense List, as described in Section 15.081 of the Texas Election Code, is maintained by the voter registrar in each county. It contains the names of (1) voters that failed to respond to a confirmation notice, (2) voters whose renewal certificate was returned to the registrar as undeliverable, and (3) those individuals that were excused or disqualified from jury service because they were not a resident of a given county.

no matches were found for 8.3 percent of people classified as Black and 5.8 percent of

people classified as Hispanic.

95.   The differences between the racial groups are statistically significant. The

difference between Black and Anglo rates of no matches is 4.0 percentage points, and the

difference between Hispanic and Anglo rates of no matches is 1.5 percentage points.

Both differences are highly unlikely to arise by chance.

Expired IDs

96. The DPS database lists records with expired IDs (DL, PID, CHL, and EIC). An

expiration may simply mean that the individual allowed the driver license to expire

because that person no longer drives.   However, an expired license may also signal that

the individual is no longer at a given residence.[32] Column 3 in Table VII.1 presents the

overall NO MATCH figure and the rate of NO MATCH among racial groups excluding

from the pool of Registered Voters those who could be matched to DPS records with

Expired IDs. It should be noted that in order to extract the expiration information, I first

matched the records in TEAM to the DPS files and then omitted from the analysis all

records that had IDs that were expired for more than 60 days and thus not valid for use

under SB 14.

---

[32] Therefore, exclusion of these records removes some registered voters who no longer
reside at the residence at which they are registered, and perhaps not in the State of Texas.
It also removes records of some people who remain at their residence but allowed their
ID to expire and will be affected by SB 14.

97. Column 3 of Table VII.1.A presents the ecological regression estimates of non-matching rates for the racial groups after removing records with expired DPS IDs. Ecological regression estimates that the NO MATCH rate among Anglos in this subset of the data was 2.1 percent. The NO MATCH among Blacks in this subset of the data was 7.6 percent. The NO MATCH among Hispanics in this subset of the data was 5.4 percent. The rates of NO MATCH among Black and Hispanic voters are statistically significantly higher than for Anglos.

98. Excluding those with expired IDs from the pool of Registered Voters, among records that the Catalist estimates classified as Anglo, 3.6 percent were not matched to any record in any identification database. By comparison, no matches were found for 5.9 percent of people identified as Black and 4.0 percent of people identified as Hispanic.

99. The differences between the racial groups are statistically significant. The difference between Black and Anglo rates of no matches and no exemptions is 2.6 percentage points, and the difference between Hispanic and Anglo rates of no matches is 0.4 percentage points. Both differences are statistically distinguishable from 0 difference.

All Filters

100. Finally, I exclude from the pool of registered voters those records with Catalist deceased, deadwood, or NCOA flags, those with Expired IDs, and those listed as Suspense Voters from TEAM. Column 4 in Table VII.1.A and VII.1B presents the

47

overall NO MATCH figure and the rate of NO MATCH among racial groups'

registrations when excluding Expired IDs, Suspense registrations, and registrations for

which there is a Catalist flag for deceased, deadwood, or NCOA.

101. Column 4 of Table VII.1.A presents the ecological regression estimates of non-

matching rates for the racial groups after removing Suspense (or inactive) voters.

Ecological regression estimates that the NO MATCH rate among Anglos in this subset of

the data was 1.9 percent.   The NO MATCH among Blacks in this subset of the data was

7.3 percent. The NO MATCH among Hispanics in this subset of the data was 5.2

percent.   The rates of NO MATCH among Black and Hispanic voters are statistically

significantly higher than for Anglos.

102.   Excluding all three categories of possible obsolete records, among those that

Catalist estimates classified as Anglo, 2.5 percent were not matched to any record in any

identification database. By comparison, no matches were found for 5.1 percent of people

classified as Black and 3.5 percent of people classified as Hispanic.   Even after

excluding these categories to test for the sensitivity of results to potential deadwood, I

find that Blacks were twice as likely to be a NO MATCH as Anglos, and Hispanics are

40 percent more likely to be a NO MATCH as Anglos.

103. The differences between the racial groups are statistically significant. The

difference between Black and Anglo rates of no matches is 2.6 percentage points, and the

difference between Hispanic and Anglo rates of no matches is 1.0 percentage points.

Both are highly unlikely to have arisen by chance.

104. These analyses suggest that the general pattern of racial differences holds even under varying constructions of the pool of registered voters. It is possible to perform further analyses using combinations of different filters and screens examined here. Appendix Tables A.VI.1 and A.VI.2 present results for the disparities on the incidence of NO MATCH for the subset of TEAM records for which Catalist shows that a record may be questionable as deceased, deadwood, or NCOA.

B. Racial Classification

105. The racial classification provided by Catalist is an estimated value for each individual based on local area demographics, frequencies of names, and other characteristics of the individuals. Errors in classification can occur. Statistical theory predicts that such errors will tend to reduce observed differences across the groups.[33] Hence, the true differences may be larger than the estimates above. It is possible to check the robustness and validity of the racial differences observed using a subset of records for which Catalist estimates of race are of the highest confidence and by using data on Spanish Surnames on the Voter files.

---

[33] William G. Cochran, "Errors of Measurement in Statistics," *Technometrics* 10 (Nov. 1968): 637-666.

106. First, the pool of registered voters can be restricted to the subset of records for which Catalist has high confidence in the racial classification. Catalist provides information on the level of confidence in that predicted value. There are 6,416,663 such records on TEAM.

107. Table VII.2 presents the numbers and percentages of each racial group in the NO MATCH and MATCH list starting with the Baseline Universe but retaining from TEAM only those records for which Catalist reports high confidence of the individual's race. Overall, 5.3 percent of records in this subset of the TEAM baseline universe did not match to a corresponding record on a state or federal identification database.

108. Racial differences in rates of NO MATCH are as more pronounced in the subset of registered voters for which Catalist reported that the racial classification estimate was "highly likely" correct.   Of people that Catalist's estimates classified as Anglo, 4.6 percent had NO MATCH. By comparison, 10.7 percent of people classified as Black and 6.2 percent of people classified as Hispanic on TEAM could not be matched to a corresponding record on any identification database.   The percent NO MATCH is 6.1 percentage points higher for Hispanics than for Anglos, and the percent NO MATCH is 1.6 percentage points higher for Blacks than for Anglos. These differences are highly unlikely to have arisen by chance and are somewhat larger than those exhibited in Tables VI.1.

50

109.   Second, the TEAM database includes an indicator of Spanish Surname Voter

Registration (SSVR).   I examine differences in NO MATCH rates for those with SSVR

and those without to gain further evidence of a racial difference.   Comparison of SSVR

with non-SSVRs will understate the differences between Hispanics and Anglos, because

the pool of non-SSVRs consists of both Anglos and Blacks.

110.   Table VII.3 presents the incidence of SSVR and non-SSVR in the NO MATCH and

MATCH list for the Baseline.   Of Spanish Surname Voter Registrations in TEAM, 6.4

percent had NO MATCH.   This estimate is slightly lower than the ecological regression

estimate for Hispanics in Table VI.1, but very similar to the estimates in Tables VI.3.A

and VI.4.A.   The estimate using SSVR is identical to the estimate of 6.4 percent NO

MATCH for Hispanics in Table VI.2 using the Catalist estimates of race, and slightly

larger than the estimates for Hispanics in Tables VI.3.B and VI.4.B.   Of those people

classified as Non-SSVR, 5.7 percent failed to match to an identification database.   The

non-SSVR pool combines Anglos, Blacks, and other races, so it does not provide a pure

comparison of Hispanics and Anglos.   Even still, I observe a difference in the rate of NO

MATCH of 0.7 percentage points between SSVR and Non-SSVR that is highly unlikely

to have arisen by chance.

VIII. Historical Voting and Registration Patterns

111.   Registration and voting patterns in prior elections can be informative about historical patterns of behavior and election administration.   To assess such historical conditions I examined three sorts of data.

112. First, I examine data from TEAM on vote history and data from Catalist on race of individuals.   These data measure whether Anglo registrants voted at higher rates than Black and Hispanic registrants in 2010 and 2012.

113. Second, I examined the CPS Registration and Voting Supplement and Census reports based on the CPS of the numbers and percentages of persons and citizens of voting age who are registered and who voted in the State of Texas and in each racial or ethnic category. I studied the reports for 2006, 2008, 2010, and 2012.

114. Third, I conducted ecological regression analyses to estimate the percentage of persons in each racial or ethnic group who are registered in the State of Texas and the percentage of registered persons in each racial or ethnic group who voted.   This methodology is an established methodology for studying voting patterns in voting rights cases; *see Thornburg v. Gingles*, 478 US 30, 52-54 (1986).

A. Catalist and TEAM

115. Tables VIII.1a and VIII.1b present the percentages of Registered Anglos, Blacks,

and Hispanics who voted in the State of Texas in 2010 and 2012, according to the

Catalist estimates of individual voter's race and TEAM data on vote history.   The table

presents all voters (active and suspense) in the top panel and only active voters in the

bottom panel. These data show that registered Anglos voted at significantly higher rates

than registered Hispanics and Blacks in the two elections immediately prior to the

implementation of SB 14.

B. Ecological Regression Estimates

116. I performed ecological regressions across VTDs to estimate the registration and

voting rates of various groups. For these estimates I used Census Enumeration counts of

the VAP, ACS estimates of CVAP, and Registration and Vote data reported by the Texas

Legislative Council.

117. I performed two sets of ecological regressions for each of elections considered, the

2006, 2008, 2010, and 2012 elections. The first set estimates the rate of Registration as a

percent of the VAP and of the CVAP for each of the groups. I regress the percent of the

VAP or CVAP that is registered on the percentages of each of the groups in the VAP or

CVAP. The second set of analyses estimates the voting rate of registered persons for

each of the groups. This is done by regressing the percent of registered persons who voted on the percentages of each of the groups in the VAP or CVAP.

118. Tables VIII.2a and VIII.2b present the Ecological Regression estimates of the percentages of adult citizen Anglos, Hispanics, and Blacks in the State of Texas who were registered in the 2006, 2008, 2010, and 2012 elections. The 95 percent confidence interval for these estimates is reported in parentheses beneath each estimated percentage.

119. According to the ecological regression estimates in Tables VIII.2a andVIII.2b, 83 to 87 percent of Anglos of Voting Age and 84 to 88 percent of Anglo Citizens of Voting Age in Texas are registered to vote.   65 to 77 percent of Blacks of Voting Age and 75 to 80 percent of the Black CVAP are registered to vote. And, 50 to 55 percent of Hispanics of Voting Age and 75 to 80 percent of Hispanic CVAP are registered to vote.   The differences between Anglo and Hispanic registration rates and between Anglo and Black registration rates are highly unlikely to have arisen by chance (i.e., are statistically significantly different from 0).

120.   Tables VIII.3a and VIII.3b present Ecological Regression estimates of the percent of registered persons who voted among specific racial or ethnic groups in Texas. According to the Ecological Regression estimates, the voting rates among registered voters of Anglo populations ranges from 10 to 20 points higher than the voting rates of Blacks, and the voting rate among registered voters of Anglo populations ranges from 15 to 30 points higher than the voting rates of Hispanic populations.

54

C. Current Population Survey

121. Table VIII.4 presents the CPS estimates of the percentages of adult citizen Anglos, Hispanics, and Blacks in the State of Texas who reported being registered to vote in the top panel and the percentages of each group who reported voting in the bottom panel. The margin of error of these estimates (a 95 percent confidence interval) is reported in parentheses beneath each estimated percentage.

122. The CPS estimates indicate that there existed no statistically significant difference between the percentages of Black and Anglo voter registration and turnout among adult citizens in the State of Texas from 2006 to 2012. The largest observed difference in the CPS estimates of registration rates of Blacks and Anglos arose in 2010 and was not statistically different from 0. However, the margin of error of these estimates is very wide, 7 to 9 percentage points for Blacks, so these data do not have much statistical power.

123. The CPS estimates indicate that Hispanic adult citizens in the State of Texas are significantly less likely to be registered to vote than Anglo adult citizens in the State of Texas. In 2012, for example, the difference in these groups' registration rates is 18.5 percentage points, with a standard error of 3.7 points. The 95 percent confidence interval for this estimated difference is 7.4 points. Hence, the difference is significantly greater than 0, even though the margin of error is quite wide.

124. The CPS estimates also indicate that Hispanic adult citizens in the State of Texas are significantly less likely to vote than Anglo adult citizens in the State of Texas. In 2012, for example, the difference in Anglo and Hispanic voting rates is 22.1 percentage points, with a standard error of 7.3 points. The 95 percent confidence interval for this estimated difference is 14.6 points. Hence, the difference is significantly greater than 0.

125. Table VIII.5 presents estimates of the percent of the Registered Persons in a given group that voted. These percentages are quite similar for Blacks and Anglos, according to the CPS figures, but the rate at which Hispanic registrants vote is estimated to be 10 to 20 points lower than for Anglos.

126. The margin of error on the CPS estimates, however, is sufficiently wide that these data support a fairly wide range of possible outcomes. Specifically, one cannot reject the hypothesis that the estimates based on the CPS in Table VIII.5 are inconsistent with the results from the Catalist data in Table VIII.1. Hence, CPS does not have the statistical power to detect differences at the level found using the NO MATCH list and Catalist racial classifications. Even so, CPS shows significant differences in registration and voting between Anglos and Hispanics.

127. Overall, the Catalist figures, CPS survey estimates, and the Ecological Regression estimates show that there are statistically significant differences across racial and ethnic groups in the rate at which individuals register to vote and the rate at which they vote, given that they were already registered. The observed differences in possession of SB 14

ID, then, only add a further potential barrier to participation for groups that already have

lower participation rates than Anglos in the State of Texas.

IX. Reply to Reports of Professors Hood and Milyo

Overview

128. Neither report discusses the substance of my two key findings. (1) There are racial differences in rates of NO MATCH according to Ecological Regression. (2) There are racial differences in rate of NO MATCH according to individual-level Catalist race estiamtes.   These results confirm and validate each other.

129.   Neither report offers any specific commentary on the ecological regressions performed using Census racial data or the individual level analyses using Catalist data.

130. My findings are consistent with evidence produced in other states, using both database matching and survey methodologies, most notably in 2012 in South Carolina, which has the advantage of a voter registration database containing both full Social Security numbers and the race of the registrant. In that case, database matching produced estimates that 3.9% of whites lacked required state or federal ID, compared to 8.3% of African-Americans and 6.7% of Hispanics.[34] My findings as to Texas here are also consistent with findings from Georgia produced by the State of Texas's own expert in the present litigation, and presented in his report.

---

[34] Charles Stewart III, "Voter ID: Who Has Them? Who Shows Them?" 66 Okla. L. Rev. 21.

131. Neither report shows that any of the objections the authors raise affect the findings

of racial disparities identified in my report. Specifically, neither report documents that

any of the criticisms offered of the matching process or of the quality of State of Texas

data alter any conclusions drawn about disparities in the rates with which Blacks and

Hispanics possess ID acceptable under SB 14 relative the rate with which Anglos possess

those IDs.

132. These reports make three broad claims regarding the conclusions in my report.

133. First, they are critical of the matching process insofar as some records on the State

of Texas official record of registered voters (TEAM database) may be obsolete

(deadwood), say because an individual is deceased or moved.   Milyo offers an estimate

that between 0 and 24 percent of records on the State of Texas voter registration list are

deadwood.   He provides no assessment as to whether his estimated deadwood records

are matched or not matched to State of Texas or federal identification databases. Nor

does he establish that any deadwood on Texas's voter registration list creates matching

problems that are correlated with race.   My original report examined exactly this

question and concluded that the racial disparities in match rates do not change with

alternative methods of identifying deadwood or deceased records through use of both

state and non-official sources.

134.   Specifically, additional analyses are performed to test the sensitivity to possible

objections concerning possible deadwood on the registration rolls.   Milyo calls the

59

"hidden results," but the results are fully reported and their relationship to the overall analysis is clearly stated in my initial report. Those analyses test the sensitivity of results excluding Suspense voters, excluding potentially exempt voters, excluding cases flagged by Catalist as deceased, deadwood, or having an NCOA flag. The disparity in the rates of NO MATCH across racial groups remain quite stable across various analyses to examine sensitivity of results to other (non-official) indicators of deadwood records and deceased voters.   Taken together, those results reveal that the racial gap shown in the data is real and durable.

135. Second, both reports are critical of the matching process. While Professor Hood lists specific problems of concern that would create a NO MATCH in the merge process, Professor Milyo argues that deadwood generally creates a problem for matching. Professor Milyo does not offer specific concerns about the matching algorithm that I used. While Professor Hood lists specific problems of concern that could create a NO MATCH in any merging process, his specific criticisms have already been anticipated and accounted for in the thorough matching algorithm that I implemented.

136. Neither report offers an alternative implementation of the matching process, even though one of the state's experts (Professor Hood) has conducted record linkage and matching for purposes of evaluating ID laws. Neither report establishes that their criticisms affect conclusions about racial differences.

137. Both Professor Hood and Professor Milyo raise concerns about the quality of the

TEAM and DPS databases maintained by the State of Texas.   I take the State of Texas

TEAM database for what it is—the official record of registered voters in the State of

Texas.   Likewise, I take the DPS data as the official record of photo IDs issued by the

State of Texas.   The contents of these databases will be used to verify whether a voter

can cast an in-person ballot.

138.   The purpose of the algorithm that I developed and implemented is to find all

individuals in TEAM who likely have a matching record in at least one identification

database, and, thus, likely have a form of ID required under SB 14.

139.   My objective is not to measure the problems with the lists maintained by the State

of Texas, and the potential barriers to voting that might arise from problems in those lists,

but to ascertain whether minority voters are going to be disparately burdened by the law

because of lower rates of SB 14 ID possession.   I perform multiple sweeps through the

relevant databases in order to guard against errors in a specific field producing a non-

match.

140. Third, the reports characterize academic literature incorrectly in three respects.

Most of literature reviewed concerns ID laws with requirements that are not comparable

to those imposed by Texas's SB 14.   Neither report explains how to extrapolate the

analyses of those laws to analysis of the effects of SB 14's requirements.

A. Response to the Report by Professor Hood

1. Academic Research on Voter ID

141.   On pages 10 and 11, Professor Hood describes studies of Voter ID laws covering

elections from 2000 to 2008, including my own 2009 paper.[35]   Most of these studies,

including my own, compare the typical ID law at the time with states that did not require

ID. Before 2008, laws requiring photo ID were atypical of ID laws.   Hence, these

studies, and the conclusions in them, do not directly apply to the present law.   No

analysis is conducted by Professor Hood to state how these studies apply to the current

law. My conclusions in my 2009 paper were based on the data presented in that paper,

and on the range of ID laws across many states with varying degrees of strictness.[36]   My

conclusions in this report are based on the data described above. My conclusions

presented here and in my 2009 paper are in no way contradictory.

142. Subsequent research has found that states that recently adopted photo ID laws have

experienced drops in turnout connected with the adoption of those laws.   An article in a

peer reviewed journal by Professors Alvarez, Bailey, and Katz in 2011 shows that the

---

[35] Stephen Ansolabehere, "Effects of Identification Requirements on Voting: Evidence
from the Experiences of Voters on Election Day," *PS: Political Science*, January, 2009,
pages 127-130, doi:10.1017/S1049096509090313.
[36] The National Conference of State Legislatures provides a summary of sates' voter
authentication rules. Wendy Underhill, "Voter Identification Requirements" National
Conference of State Legislatures, June 25, 2014, http://www.ncsl.org/research/elections-
and-campaigns/voter-id.aspx

states that introduced photo ID had a significant drop in turnout, and that the photo ID requirements had qualitatively different effects on turnout than other laws.[37]   Other, less restrictive ID laws had minimal effects on turnout in their analysis, but requiring photo ID corresponded with a significant decline in turnout.[38]

143.   Professors Hood and Bullock in their 2008 study of the Georgia photo ID law establish the link between lack of ID and decreased turnout.   That study uses record linkage and database matching between the voter registration list in Georgia and the driver license list in Georgia. The authors found that "[r]egistrants who lack drivers licenses are generally less engaged politically and maybe even less apt to participate if more ID restrictions are put in place. (Page 573)." See also their 2012 article.[39]

144.   Professors Hood and Bullock further established the link between race and lack of ID in Georgia. They concluded that "[r]egistered voters are significantly less likely to possess a drivers license if they are from minority groups, especially if they are Black and Hispanic, and if they are older." (Page 572)   According to Hood and Bullock, then, Blacks and Hispanics are significantly less likely than Whites to possess drivers licenses, and those without drivers licenses are significantly less likely to vote following the implementation of the Georgia photo ID law.

---

[37] R. Michael Alvarez, Delia Bailey, and Jonathan Katz, "An Empirical Bayes Approach to Estimating Ordinal Treatment Effects, *Political Analysis* 19: 20-31.

[38] Mycoff, Wagner, and Wilson, "The Empirical Effects of Voter-ID Laws: Present or Absent?" PS: Political Science January, 2009, pages 121-126, examine CCES survey data and find a negative effect of photo ID requirements, but the statistical precision is insufficient to determine whether the true effect is statistically distinguishable from 0.

[39] M. V. Hood and Charles S. Bullock III, "Much Ado About Nothing? An Empirical Assessment of the Georgia Voter Identification

145.   In his report, Professor Hood characterizes his study as showing no racial

difference in turnout with the introduction of the photo ID law in Georgia.   This

conclusion arises from the analyses in Tables 2 and 3 in his report.   Those analyses

estimate the differential drop in turnout between those who had ID and those who did not

have ID from 2004 to 2008 in Georgia across racial groups.   He concludes that there is

no difference across racial groups in the change in differential turnout.   Professor Hood's

interpretation and methodology in reaching that conclusion are flawed.

146. The interpretation ignores the fact that there are racial disparities in possession of

ID.   The analysis in Tables 2 and 3 addresses the decline in turnout within groups for

those who possessed and did not possess IDs in Georgia. It does not address the question

of the total effect of the ID law given differential rates with which groups possess ID.   A

hypothetical example demonstrates the point.   Suppose there are two groups.   In one

group 98 people have ID and 2 do not.   And in a second group, 2 have ID and 98 do not.

Before the ID law is implemented all people with ID vote and half of those without ID do

not.   Once the law is implemented all people with ID vote and none of those without ID

vote.   The decline in turnout is the same within both groups. That is, 100 percent of

those with ID voted before the law and 50 percent of those without ID voted before the

law within both groups, and the decline in voting from before to after is the same.

However, the effect of the law falls almost entirely on the second group because of the

lower rate of possession of ID.   Specifically, the turnout rate of the first group would be

99 percent before the law and 98 percent after the law, but the turnout rate of the second

group would be 51 percent before the law and 2 percent after the law.[40]   This

hypothetical assumes that the differential decline in turnout is the same in both groups,

and it demonstrates that that is not the measure of the total effect of the law on turnout of

groups.

147.   Other scholars have issued methodological criticisms of the study of the Georgia

photo ID law by Professors Hood and Bullock.   Professor Stewart provides a thorough

assessment of some of the weaknesses of the research design, with which I agree. (See

attached report of Professor Charles Stewart in *South Carolina v. United States*.)   Two

critiques are especially important in ascertaining whether there is an effect of photo ID

laws.

148.   First, differences in income and other demographics across groups will translate

into differential effects of photo ID laws on minorities. Drawing on the results reported

in his research, Professor Hood reports results of multivariate logit analyses, which hold

constant income, gender, age, and area of residency. He concludes that there is no racial

effect of photo ID laws on turnout. However, the coefficient on race in the logistic

regression is not the average effect of the law on a typical Black person, a typical

Hispanic person, or a typical White person.   It is the difference across racial groups,

assuming all racial groups are the same in all other demographics, including income.

---

[40] Before the law 1 of the 2 people in group 1 without ID would vote and all 98 of those
with ID would vote, for a turnout rate of 99 percent.   After the law, all 98 with ID would
vote and neither of those with ID would vote.   For the second group, before the law, 49
of the 98 people (half) without ID would vote and both of those with ID would vote, for a
turnout rate of 51 percent.   After the law, all 98 without ID would not vote and both of
those with ID would vote.

That is, there is no significant difference in expected turnout between a white and a black individual who have the same income, gender, age and area of residency.

149.   All things, of course, are not equal. Blacks and Hispanics have lower income and are younger than Whites in Georgia[41] and in Texas.[42]   The introduction of statistical controls makes it appear in the multivariate logit analysis that there are no racial differences when in fact there may be one because minorities have lower values on average on the relevant statistical controls.

150.   Second, Professor Hood's estimate of the decline in turnout among those without ID is biased and biased in a way that will overstate the decline in White turnout from 2004 to 2008 relative to the decline in Black and Hispanic turnout. The comparison of 2004 and 2008 in Hood and Bullock, and reproduced in Tables 2 and 3 of Professor Hood's report, rests of knowing who had Driver Licenses in 2004 and in 2008.   To determine that, Professors Hood and Bullock engage in record linkage and database matching of the Georgia voter file to the Driver License list. The databases are already linked through a unique identifier. The list of individuals who do not have driver licenses was produced late in 2007.

---

[41] See Rebuttal Declaration of Professor Charles Stewart, in South Carolina versus United States, United States District Court of District of Columbia, Document 166-2, especially pages 38-53.
[42] See Granted Request for Judicial Notice, in *Veasey v. Perry*, in the United States District Court for the Southern District of Texas, Corpus Christi Division, Document 252

151.   The matching procedures used by Professor Hood had dramatically different rates of matching for 2004 and 2008. Professors Hood and Bullock relied solely on the "unique identifier," and they matched 98.2% of 2008 registered voters to the 2007 list of those without driver license, but only 78.5% of 2004 registered voters to the 2007 list of those without drivers licenses. It is improbable that there was a 20-percentage point increase in the number of registered voters who actually had drivers licenses between 2004 and 2008. More likely the matching algorithm did substantially worse in that year, making the estimates for 2004 much worse than for 2008 and introducing a roughly 20-point bias in any direct comparison of ID possession between the two years.

152.   The 98.2 percent matching rate that Professors Hood and Bullock found through use of a unique identifier in record linkage is approximately the same level of reliability that the algorithm I implemented attains using multiple identifiers that combine Address, Name, Gender, and Data of Birth. (See the comparison with SSN9 matches below.)   The reliability of the match rates of the algorithm implemented for the United States in the present case far exceeds the 78.5% rate that Hood and Bullock found as to the 2004 registered voters.

153.   The large difference between the 2004 and 2008 match rates in the analysis performed by Professors Hood and Bullock will bias the estimates of the drop in turnout and, if there are racial differences in possession of ID, will bias the estimated effect of the law on those who do no possess ID across racial groups.   The bias in the drop in turnout arises because many people who actually have ID are classified as not having ID in 2004,

but not in 2008.   Since those with ID vote at a much higher rate than those without ID

(even before the new photo ID law went into effect), those classified as not having ID in

2004 are a mix of those without ID, who vote at a low rate, and some people with ID,

who vote at a higher rate.   That will inflate the estimated voting rate of those who do not

in fact have ID.

154.   If possession of ID is correlated with race, the misclassification will inflate the

estimated turnout rate of Whites without ID more than it will inflate the estimated turnout

rate of Blacks and Hispanics without ID.

155.   In sum, the methodological problems with the Hood and Bullock estimates (and

with Tables 2 and 3 in Professor Hood's report) caution against drawing reliable

conclusions about the racial disparities in the effects of the introduction of photo ID rules

in Georgia on turnout.   Methodological problems aside, the results in Tables 2 and 3 do

not address the fact of racial disparities in rates of possession of IDs; hence, the results

cannot be interpreted as the total effect of the introduction of the photo ID laws on voter

participation by race.

2. Matching Algorithm

156. Professor Hood offers six specific criticisms of my matching algorithm.   I respond

to each in kind.

68

a. There is no universal identifier in these databases.

157. The State of Texas does not record a unique identifier, such as SSN9, for every record on that corresponds to an individual on DPS and other state databases.   Not every TEAM record has an SSN9 or DPS ID number.

158. As Hood and Bullock's work shows even with such a "universal identifier" subsequent matching procedures are required.   The universal identifier in their study of Georgia matched 98.2 percent of records in 2008 and 78.5 percent for 2004.

159. The state of science is to use multiple identifiers, either using exact matches or probabilistic matches, even when there is a universal identifier because of problems of missing data and typographical errors.   That is what I have done.   The reliability of this method is further shown through the analysis of possible false positive and false negative matches below.

b. Missing data in fields

160. As I discussed in sections IV.B and IV.C, missing data in fields is one reason for using multiple identifiers for record linkage. Even the universal identifiers have missing data (and other errors) such that use of multiple identifiers is preferable.

161. The use of multiple identifiers avoids non-matching arising from a field being missing in one database.   For example, one matching combination I used includes only Address, Date of Birth, and Gender.   Because there are no Name elements in this identifier, missingness in name will not make a match impossible.

162.   Using multiple indicators avoids non-matches arising from a field being missing in one database but not in other databases.   If one field is missing, no identifier is created for purposes of matching for any indicator that relies on that field. If that occurs in TEAM no match is attempted for that identifier.   If that occurs in an identification database, no match occurs for that record. However, we match on multiple identifiers and there is always at least one identifier that does not rely on a given field.   For example, the combination of Address, Gender, Date of Birth that I use does not rely on Name, so missingness in name does not mean a match is impossible.

c. Inconsistencies in Fields between Databases

163. Professor Hood states that inconsistencies in fields across databases can lead to incorrect NO MATCH. On page 20 of his report, Professor Hood provides the example of a person who is recorded with the name JIM SMITH in one file and JAMES SMITH and the same DATE OF BIRTH, ADDRESS, and GENDER on both files.

164.   The algorithm used was developed to avoid this and similar problems. The variations built into the matching combinations avoid non-matches that would arise from

70

a wide variety of issues, such as variations in first names and nicknames, hyphenated or multiple last names, variations in recording of addresses.

165.   In the particular example Professor Hood offers, the algorithm would have found a match because certain matching combinations do not include first name, while others do not include any name element at all.

166.   Professor Hood offers no other examples of potential problems, nor any assessment of the likely frequency of such problems.

d. Errors in Fields

167. Errors in fields might also create inconsistencies between analogous fields in databases. Here again, the use of multiple indicators ensures that typographical and other errors in fields do not prevent matches from occurring. A typographical error in, say, ZIP code, would not prevent a match on Date of Birth, Gender, and Name.

168. Professor Hood provides the example of the high frequency of birthdates on the first date of the month or on November 11. But identifiers that do not use Date of Birth, such as Address, Gender, and Name will still link match these records, even when there are errors in Dates of Birth.

e. Deceased Fields Do Not Identify All Deceased Persons

169. Our approach to deceased individuals is to take the State of Texas databases at face value, and then to conduct additional analyses using Catalist identifiers for the likely deceased.   After matching has occurred, I use all DPS indicators for deceased persons to remove from the set of TEAM records to be analyzed all individuals matched to records DPS has marked as deceased.

170. In additional analyses, we further remove individuals that Catalist identifies as likely deceased.   These additional analyses test the sensitivity of results to Deceased persons and other forms of obsolete records, and they show that deceased records do not affect inferences concerning racial differences in rates with which NO MATCH is found between TEAM and identification databases.

171.   Professor Hood offers no evidence deceased individuals remain on the rolls after our matching process or after the incorporation of information from Catalist, nor that any such records bias the racial disparities found.

f. No Race or Ethnicity on Database

172. The fact that race is not on the voter files does not preclude analyses of racial disparities in election laws. I approach this matter two ways: using ecological regression and Census data on race and using classification of race of individual records on TEAM provided by Catalist. Ecological Regression analysis is a well established and accepted

methodology for using aggregate election and racial data, say at the level of block groups or voting tabulation districts, to estimate the behavior of various racial groups.   In *Thornburg v. Gingles* 478 US 30 (1986), the Supreme Court established that ecological regression is an acceptable form of evidence in questions arising under section 2 of the Voting Rights Act. Ecological regression using Census racial data shows statistically significant differences in rates of NO MATCH among racial groups.

173. Race is on the voter files in only 9 states, but not Texas.   In lieu of individuals' self-reported race, I analyze data on the likely race of individuals provided by Catalist, LLC.   Analysis of that information shows significant differences in the rates of NO MATCH among racial groups, consistent with the conclusions of Ecological Regression analyses.

174.   The State of Texas makes its own race-related variable, in Spanish Surname Voter Registration (SSVR).   That indicator is on the voter files. We analyze SSVR versus other records and find a racial disparity between those identified as Hispanic (Spanish Surname) are less likely to have a matching record on an identification database than are other registered voters.   Also the NO MATCH rate for SSVR is quite similar to that found using the analysis of Census racial data and Catalist racial data.

g. No Post Estimation Validation

175. Professor Hood states that I reported no analysis of potential false positives and false negatives in the matching process. He does not specify what exact analysis ought to be done, nor does he offer an analysis and evidence indicating why he thinks there may be an unusual number of false positives and false negatives.

176.   In developing the algorithm for the primary matches, I did perform post estimation validation. I used matches to SSN9 as a "unique identifier" against which to test the accuracy of the primary matches using combinations of Address, Date of Birth, Gender, and Name. For the subset of records with SSN9 on TEAM (approximately half of the records), I examined cases that had SSN9 and for which there was NO MATCH between TEAM and a DPS record using combinations of Address, Date of Birth, Gender, and Name. That is, to test the validity of the Primary Matching algorithm, I conducted those matches for cases with SSN9. I then rematched the cases to DPS using SSN9, and calculated the percent of cases for which no Primary Match could be found but for which there was an SSN9 match.

177. In preliminary analyses approximately 5 percent of these NO MATCH records had a match to DPS on SSN9.   Matching on SSN9 (approximately half of cases), matching on Texas DL (a primary match but not used for the validation), and matching on secondary matches using combinations of Name, Address, and SSN4 would reduce that number further.

74

178. I conducted this analysis again with the complete TEAM and DPS data on August 7, 2014. Again, I restrict the analysis to the subset of records on TEAM that have an SSN9 and for which validation of the matching algorithm is possible.   First I examined the set of records for which there was a match using SSN9 and ascertained how many failed to match to have a primary match using Address, Date of Birth, Gender and Name. Of the records with 5,041,166 records that match on SSN9 between TEAM and DPS, 2.6 percent were not matched using Address, Date of Birth, Gender and Name between TEAM and DPS.   I further examined the set of cases for which there was a primary match using Address, Date of Birth, Gender, and Name.   Of the 5,028,183 records on TEAM that match to one of the primary indicators using Address, Date of Birth, Gender, and Name between TEAM and DPS, 2.6% do not have a matching SSN9 between TEAM and DPS. Hence, the Address, Date of Birth, Gender, and Name combinations could accurately match 97.4 percent of records (using SSN9 as the benchmark for validation). By comparison, SSN9, which is often relied on as a unique identifier, could match 97.6 percent of records (using Address, Date of Birth, Gender, and Name primary matches as a benchmark for validation.   In other words, the primary matches on combinations of Address, Date of Birth, Gender, and Name are almost the functional equivalent to matching on SSN9.   See Table X.1.

179. The rate at which the primary sweeps using Address, Date of Birth, Gender, and Name combinations yield NO MATCH but for which a matching SSN9 exists is extremely low.   It is lower, for example, comparable to the rates reported in the article of

Professors Hood and Bullock on the Georgia ID law for 2008 (98.2%), and much lower than the figure they report for 2004 (78.5%).

180. The rate at which the primary sweeps using Address, Date of Birth, Gender, and Name combinations yield NO MATCH but for which a matching SSN9 exists is lowered further upon using DPS ID in the primary matches, upon conducting secondary matches, including SSN9, and upon using the federal data.   It is also worth noting that not all of the cases in this small subset of NO MATCHES are erroneous. Some of these NO MATCHES may reflect discrepant information on TEAM and DPS that would make it difficult to authenticate the person at the voting place, such as a name change and missing Date of Birth, or a change in name and address.

3. Analyses of the No Match List

a. NO MATCH records with State IDs

181. Professor Hood presents an estimate of the number and fraction of cases on the NO MATCH list with a DPS ID.   This is meant to suggest that these are cases that should have been matched, but were not.   Professor Hood does not provide evidence that any of these records were in fact found on the DPS lists and had valid ID.

182. A breakdown of the NO MATCH list reveals that 90 percent of these cases affirmatively do not have a valid SB 14 ID. The remaining 10 percent could not be found on the DPS list.

183. Of the 786,727 records on the NO MATCH list, 446,180 have a DPS ID number (e.g., a driver license or state ID) and 340,551 have no DPS ID number.   That is 56.7% have a DPS ID number.

184. Of the 446,180 with a DPS ID, 226,054 records have an expired ID and 220,126 have a non-expired ID.

185. Of the 220,126 on the NO MATCH list who have a non-expired ID, 138,081 were originally on the MATCHED using the January 2014 list but were determined after receiving additional information provided by the State of Texas in July to have a card status or license surrender status indicating that the individual does not currently posses the driver license or personal ID associated with that particular record.

186. The remaining 82,045 records had a DPS ID number on TEAM and were classified as NO MATCH. I searched for the DPS ID number and the SSN9 (for those that had one) in the DPS file and determined that none of these cases match to the DPS file using the DPS ID, SSN9, or on any sweep using the algorithm I developed.   I performed a visual inspection of the databases of 100 randomly selected cases from these lists for similar names and other identifying information and did not find any similar records.

187. There are several reasons why these people would have a DPS ID and not be on the DPS. A non-exhaustive list includes the following: (1) The underlying DPS record may have been purged, perhaps because they were no longer valid and were removed as deadwood from DPS. (2) The DPS records provided to me may not be complete.   There might, for example, be a programming problem in the creation of the file using DPS, as occurred in January. (3) There might be typographical errors in the DL number (though, if that alone were the cause, the other matching sweeps likely would have caught some of these records).

188. Hence, of the 786,727 persons on the NO MATCH list, 340,551 (45.3%) do not have a record of a DPS ID on TEAM. Of those that do have a DPS ID on TEAM, 220,126 have an expired DPS ID, 138,081 have a problem with the DPS ID such as a surrendered license, and 82,045 have a DPS ID according to TEAM but are not evidently on the DPS file.

189. These records, then, appear to be genuine records of NO MATCH because they have no ID, have no valid ID, or are not included in the DPS dataset.

b. 2013 and 2014 Turnout

190. Professor Hood analyzes turnout data for 2013 and 2014 among the records on the

NO MATCH list. Approximately 7 percent of records on the NO MATCH list using the

January 2014 data were determined to have voted in an election in 2013 or 2014.

191.   I examined the number and percent of records on the NO MATCH list that voted

in the 2014 Primary and Primary-Runoff election, after re-doing the matching algorithm

with the supplemental data provided by the State of Texas.   Of the 786,727 records on

the NO MATCH file (after removing deceased records), 27,769 (3.5 percent) are

recorded as having voted in the 2014 Primary and 12,994 (1.7 percent) are recorded as

having voted in the 2014 Run-off election. 22,387 (2.8 percent of the NO MATCH)

voted in person rather than absentee.

192. There are a variety of explanations that are possible for these cases. Including:

Individuals may have updated their registration records between January and the election

dates; they may have obtained a new license in that interval; they may have had

unmatchable records; poll workers may have let these individuals vote without the

required ID.   20 percent voted absentee by mail.

B. Response to the Report by Professor Milyo

193. Professor Milyo begins his discussion of my analysis of record linkage and

matching of the State of Texas voter registration records with an assessment of the

79

existing research on the subject. In paragraph 21 he states that he is "unaware of any

scholarly studies that analyze the effects of voter ID by examining 'non-matches'

between a state voter registration database and external databases." (See also paragraph

22.) This statement is incorrect. The article by Professors Hood and Bullock, "Much

Ado About Nothing? An Empirical Assessment of the Georgia Voter Identification

Statute," performs exactly such a record linkage and matching process in order to

determine the number of non-matching records between the voter registration records and

the Drivers License files in the state of Georgia.[43]

194. Professor Milyo's specific critiques of the matching process are summarized in

paragraph 22 of his report:

"It is well known that state registration databases contain errors and that in general counts

from such databases exaggerate the actual number of eligible and currently registered

voters." (Paragraph 22) He references my research on this subject, which relies on

database errors identified by Catalist.   And as support for his conjecture states that "[i]t

is for this reason that much of the scholarly research on voter turnout in the United States

eschews measures of turnout as a percent of registered voters and instead examines

turnout relative to voting age population (VAP), citizen voting age population (CVAP),

or eligible voting age population (VEP)."

---

[43] Professor Milyo's report references this as *Legislative Studies Quarterly* in the citation
in footnote 112 on page 33. The article appears in *State Politics and Policy Quarterly* 12
(2012): 394-414.

195. As discussed in the Overview and below, the analyses performed examine whether

conclusions drawn using only the official State of Texas and federal databases to

determine valid registrations are altered once potentially invalid cases, as identified by

other data sources, are removed.   In particular, one set of identifiers of potential

deadwood, movers, and deceased individuals comes from Catalist, and is the source of

the assessment of problematic records on registrations. My report and the analyses in the

preceding part of this supplemental report show that the conclusions about the relative

racial disparities in rates of non-matching are not sensitive to indicators of deadwood.

196. Professor Milyo states that political scientists "eschew measures of turnout as a

percent of registered voters." This is is untrue. Registration lists are widely used in

research on turnout, and turnout as a percent of registrations is a commonly used measure

of participation rates.   A sampling of applications of the use of registration lists for

participation research includes: measures of turnout rates overall and of minority groups,

in research nationwide and on the state of Texas,[44] turnout rates in local areas,[45] effects

of election laws,[46] effects of political campaigns on participation,[47] and as the official list

of registered persons for purposes of drawing random sample surveys. On the last point,

---

[44] Daron Shaw, Rudolfo O. de la Garza, and Jongho Lee, "Examining Latino Turnout in
1996: A Three-State, Validated Survey Approach," American Journal of Political
Science 44 (2000):   338-346.
[45] James G. Gimpel, Joshua J. Dyck, Daron R. Shaw, "Registrants, Voters, and Turnout
Variability Across Neighborhoods," Political Behavior 26 (2004): 343. See esp. page
348.
[46] Priscilla L. Southwell and Justin Burchett, "The Effect of All-mail Elections on Voter
Turnout," American Politics Research 28 (2000): 79-79
[47] Alan S. Gerber, James G. Gimpel, Donald P. Green, and Daron R. Shaw, "How Large
and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a
Randomized Field Experiment" American Political Science Review 105 (2011): 135-
150.

research by Professors Donald Green and Alan Gerber finds that registration based

sampling provides more accurate forecasts of future election results than does traditional

sampling procedures.[48]

197. Professor Milyo's central concern is with the amount of "deadwood" on the

registration lists.   He offers no direct examination of actual voter records to determine

the number of such records. In paragraphs 23 to 25, he calculates the rate of deadwood

two ways. First, he compares the registration rate for the state of Texas based on the

number of records on the voter registration lists divided by the American Community

Survey (ACS) estimate of the Citizen Voting Age Population with the Current Population

Survey (CPS) estimate of the percent of voting age citizens who are registered as reported

in the November supplement of the CPS. Second, he assumes over reporting of

registration and deflates the CPS registration number and inflates his estimate of

deadwood.

198. In paragraph 23 he calculates that there is an 8.3 percentage point difference

between the official records and the CPS reports (81.2 percent versus 72.9 percent).

199. He offers no statistical assessment, or standard error with this figure, even though

the CPS is a random sample of a few thousand people of each state's population and,

---

[48] Donald R. Green and Alan S. Gerber, "Can Registration-Based Sampling Improve the Accuracy of Midterm Election Forecasts?" *Public Opinion Quarterly* 70 (2006): 197-223.

thus, is estimated with uncertainty.   The CPS figure, then, is itself an uncertain number.

No margin of error is reported with his projected number of Deadwood.

200. In making his assessment of the amount of deadwood, he makes no assessment of

which records are likely Deadwood or offers remedies for questionable records on the

registration files so that a correct assessment the matching algorithm can be conducted

201. In paragraph 25 he states that there may be reporting problems with the CPS. This

claim, of course, weakens the conjecture that the CPS should be taken as the measure of

actual registration or turnout. Further, he assumes that those reporting problems take the

form of over estimation of actual registration. No statistical evidence of the actual

amount of over reporting in CPS is presented.

202. Professor Milyo offers a range of potential deadwood from 0 to 24 percent. This is

a very wide interval for an estimate, and suggests that this methodology for measuring

the amount of deadwood is not very precise. Also, no standard errors (measures of

uncertainty of the estimate) are presented.

203. Taking the interval 0 to 24 as an interval estimate of the possible true underlying

rate of deadwood, his methodology does not appear to reject the hypothesis that the true

rate of deadwood is 0. Zero is in the interval of potential values estimated using his

methodology. It may be that Professor Milyo does not intend to construct such an

interval estimate using his methodology, but, if that is the case, it is unclear what statistical inferences can be drawn from the analyses presented.

204. Professor Milyo's calculation is uninformative about the rates of NO MATCHES. It is simply an estimate of deadwood. It has no information about the rate with which records on TEAM match to records on state and federal identification databases.   It is possible that all of his potential deadwood cases in fact MATCH to an identification database.

205.   Professor Milyo claims in paragraph 15 on page 3 that the racial effects are biased upward in my analysis resulting from database matches, but he provides no evidence to that effect. Specifically, he provides no estimate of the amount of Deadwood or resulting biases by racial group.

206. My analysis of the TEAM data is sensitive to the possibility that some of the records are obsolete, individuals who have moved, or deceased.   The database process, as described, removes all individuals identified as deceased using the information on the State of Texas DPS files.   Again, as stated in my original report, I take the State of Texas databases as the official record of registered voters and holders of identifications.   To examine whether the inferences drawn are sensitive to potential deadwood, I impose various filters for potential deadwood, using indicators from TEAM, DPS, and Catalist. These are: Suspense Voters on TEAM; Expired IDs on DPS; Catalist indicators of Deadwood, Deceased, and NCOA from Catalist. These indicators may be over inclusive

84

in classifying deadwood. For example, some registrants on the Suspense list vote in the

next election. So analyses with these indicators erred on the side of excluding too many

cases as deadwood rather than too few.

X. Conclusion

207. This analysis has found statistically significant and robust racial differences in the rate with which registered voters in the TEAM database fail to match to records in state and Federal databases of people with photo identification required for voting under SB 14. The results in Table VI.1 show that 4 percent of Anglos DO NOT MATCH to applicable SB 14 identification databases, compared with 12 percent of Blacks and 9 percent of Hispanics.   The results are strikingly similar in analyses relating Census racial data to the incidence of NO MATCH and in analyses relating individual level data on NO MATCH and to Catalist estimates of individual race. Those differences persist when I consider eligibility for exemption, alternative definitions of the pool of registered voters, and alternative racial classifications (such as SSVR). The observed differences imply that Black and Hispanic registered voters are significantly less likely than Anglo registered voters to possess applicable SB 14 ID or qualify for an exemption under the law.

I declare under penalty of perjury that the foregoing is true and correct. Executed
this 15th day of August, 2014

_____
Stephen D. Ansolabehere

TABLES

Table V.1. Combinations of Fields Used as Matching Identifiers	
Combination Code	PRIMARY MATCHES
A	First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number
B	Last Name + Gender + DOB + Residential ZIP + Residential Street Number
C	Gender + DOB + Residential ZIP + Residential Street Number
D	First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number
E	First Name + Last Name + Gender + Residential ZIP + Residential Street Number
F	First Name + Last Name + Gender + DOB
M	Texas Driver License Number (where available)
	SECONDARY MATCHES
G	First Name + Middle Initial + Last Name + DOB
H	Last 4-Digit SSN + DOB + Residential ZIP
I	Last 4-Digit SSN + First Name + Last Name + DOB
K	First Name + Last Name 1 + Middle Initial + DOB
L	First Name + Last Name 2 + Middle Initial + DOB
SSN	9-Digit Social Security Number

		Matching Combinations
Texas DPS Databases	Primary Sweeps (All TEAM records)	Combination A: First name + Last name + Gender + DOB + Street number + ZIP Combination B: Last name + Gender + DOB + Street number + ZIP Combination C: Gender + DOB+ Street number + ZIP Combination D: First name + Last name + Street number + ZIP Combination E: First name + Last name + Gender + Street number + ZIP Combination F: First name + Last name + DOB + Gender Combination M: Texas Driver License Number
	Secondary Sweeps (TEAM records with no primary match)	Combination G: First name + Last name + Middle Initial + DOB Combination H: DOB + ZIP + SSN4 Combination I: First name + Last name + DOB + SSN4 Combination K: First name + Last name 1 + Middle Initial + DOB Combination L: First name + Last name 2 + Middle Initial + DOB49 SSN: 9-digit Social Security Number
Federal Identification and Disability Databases	Primary Sweeps (All TEAM records against Federal records with a Texas address)	Same as primary sweeps for DPS databases, except for Texas Driver License Number (Combinations A-F)
	Secondary Sweeps (TEAM records with no primary match against Federal records with a Texas address)	Same as secondary sweeps for DPS databases (Combinations G-L and SSN)
	Nationwide Sweeps (TEAM records with no primary or secondary match against nationwide Federal records)	All sweeps without address criteria (Combinations F, G, I, K, L, and SSN)

---

[49] "Last name 1" is the first half of a hyphenated last name, and "Last name 2" is the second half of a hyphenated last name. Combinations K and L in TEAM are each matched against Combination G, Combination K, and Combination L in the identification and disability databases for a total of six matching sweeps.

Table V.2. Number of Matches of TEAM Records to State and Federal Databases Overall and By Racial Group, Before Removing Records Indicated as Deceaseed by DPS, using Catalist Racial Estimates

(Percent of TEAM Records that Match to a Given ID or Disability Database)

Database	Race				
State of Texas ID Databases	White	Black	Hispanic	Other	All
Driver License	7,201,903 (86.8%)	1,283,720 (74.7%)	2,455,518 (80.3%)	423,139 (85.8%)	11,364,280 (83.8%)
Personal ID	417,122 (5.0%)	312,017 (18.1%)	495,983 (16.2%)	28,666 (5.8%)	1,253,788 (9.2%)
Concealed Handgun License	588,087 (7.1%)	57,129 (3.3%)	72,953 (2.4%)	14,839 (3.0%)	733,008 (5.4%)
EIC	69	43	51	0	163
Federal ID Databases					
DOS	3,776,207 (45.5%)	424,682 (24.7%)	1,151,608 (37.7%)	378,666 (76.8%)	5,731,163 (42.3%)
DOD	427,191 (5.2%)	81,688 (4.8%)	116,460 (3.8%)	13,015 (2.6%)	638,354 (4.7%)
USCIS	106,051 (1.3%)	45,005 (2.6%)	373,576 (12.2%)	210,454 (42.7%)	735,086 (5.4%)
VHA (VIC)	186,695 (2.3%)	49,179 (2.9%)	57,635 (1.9%)	2,496 (0.5%)	296,005 (2.2%)
Federal Disability Databases					
SSA: Disability	419,065 (5.1%)	167,980 (9.8%)	202,368 (6.6%)	14,925 (3.0%)	804,338 (5.9%)
VBA: Disability	118,883 (1.4%)	31,952 (1.9%)	35,743 (1.2%)	1,938 (0.4%)	188,516 (1.4%)

89

Table V.3.A Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database using DOJ Algorithm*
(includes cases DPS flags as deceased)

		ANY FEDERAL RECORD		
		No Match to a Federal ID	Match to a Federal ID	ALL
ANY STATE RECORD	No Match to State ID	801,029	600,209	1,401,238
	Match to State ID	6,437,247	5,725,913	12,163,160
	ALL	7,238,276	6,326,122	13,564,398

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table V.3.B Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database using DOJ Algorithm*, Excluding cases DPS flags as deceased

		ANY FEDERAL RECORD		
		No Match to a Federal ID	Match to a Federal ID	ALL
ANY STATE RECORD	No Match to State ID	786,727	597,227	1,383,953
	Match to State ID	6,395,667	5,707,956	12,103,623
	ALL	7,182,394	6,305,182	13,487,576

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table V.4.A Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database or any Federal Disability Database using DOJ Algorithm* (includes cases DPS flags as deceased)				
		ANY FEDERAL RECORD		
		No Match to a Federal ID or Disability	Match to a Federal ID or Disability	ALL
ANY STATE RECORD	No Match to State ID	715,804	685,434	1,401,238
	Match to State ID	5,934,910	6,228,250	12,163,160
	ALL	6,650,714	6,913,684	13,564,398

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table V.4.B Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database or any Federal Disability Database using DOJ Algorithm* Excluding cases DPS flags as deceased				
		ANY FEDERAL RECORD		
		No Match to a Federal ID or Disability	Match to a Federal ID or Disability	ALL
ANY STATE RECORD	No Match to State ID	701,696	682,257	1,383,953
	Match to State ID	5,895,818	6,208,501	12,103,623
	ALL	6,596,818	6,890,758	13,487,576

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table VI.1.   Estimated Percent No Match By Racial Group Using Census Racial Data: Ecological Regression Analyses of ACS CVAP and No Match Percent at Block-Group Level

	Ecological Regression*	Homogeneous Block Groups***
Racial Group	Estimated % No Match (Margin of Error)	Estimated % No Match (Margin of Error)
Anglo	3.6% (± 0.1%)	4.5% (± 0.2%) [N of Block Groups = 4,224]
Black	11.9% (± .2%)	12.3% (± 0.4%) [N of Block Groups = 465]
Hispanic	8.6% (± .2%)	8.8% (± 0.4%) [N of Block Groups = 1,554]
	Gross Percentage Point Disparity in Rate of NO MATCH	
Black % - Anglo %	8.3%	7.8%
Hispanic % – Anglo %	5.0%	4.3%
	Percent Difference in Rate of NO MATCH	
(Black %-Anglo %)/ Anglo %	231%	273%
(Hispanic %-Anglo %)/ Anglo %	139%	196%

* Number of Cases = 15,673 R2 = .314.
** Level of analysis: Block Group;
Dependent variable: Number NO MATCH in Block Group divided by ACS CVAP Estimate in Block Group;
Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent; Weighted by CVAP.
*** Homogeneous block groups are areas in which at least 80 percent of the CVAP is of a given population.

Table VI.2.   NO-MATCH and MATCH Percent By Racial Group, Using Catalist Racial Classification*

Race	NO-MATCH	MATCH	ALL
Anglo	422,500 (5.1%)	7,823,516 (94.9%)	8,246,016
Black	157,811 (9.2%)	1,549,958 (90.8%)	1,707,769
Hispanic	194,204 (6.4%)	2,848,293 (93.6%)	3,042,497
Other	12,212 (2.5%)	479,100 (97.5%)	491,312
All	786,727 (5.8%)	12,700,867 (94.2%)	13,487,594
	Gross Percentage Point Disparity		
Black% – Anglo%	4.1		
Hispanic% – Anglo%	1.3		
	Percent Difference in Rate of NO MATCH		
(Black%-Anglo%) /Anglo%	80%		
(Hispanic% - Anglo %) /Anglo%	25%		

* Baseline Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.3.A. RATE OF NO-MATCH/NOT EXEMPTION ELIGIBLE/Not Age Vote-by-Mail Eligible			
By Racial Group, Using Ecological Regression*			
RACE	NO MATCH / Not Exemption Eligible	NO MATCH / Not Age Vote-By-Mail Eligible	NO-MATCH/ NOT EXEMPTION ELIGIBLE/ Not Age Vote-by-Mail Eligible
Anglo	3.3% (+/- 0.1)	2.6% (+/- 0.04)	2.4% (+/- 0.05)
Black	9.8% (+/- 0.2)	9.3% (+/- 0.2)	7.7% (+/- 0.1)
Hispanic	7.6% (+/- 0.2)	6.6% (+/- 0.1)	5.9% (+/- 0.1)
N	15,673	15,673	15,673
R-Squared	.302	.332	.308
	Gross Percentage Point Disparity		
Black% – Anglo%	6.5%	6.7%	5.4%
Hispanic% – Anglo%	4.3%	4.0%	3.5%
	Percent Difference in Rate of NO MATCH		
(Black%-Anglo%) /Anglo%	197%	258%	221%
(Hispanic% - Anglo %) /Anglo%	130%	154%	146%

* Baseline Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.3.B. NO-MATCH/NOT EXEMPTION ELIGIBLE/Not Age Vote-by-Mail Eligible By Racial Group, Using Catalist Racial Classification*			
RACE	NO MATCH / Not Exemption Eligible	NO MATCH / Not Age Vote-By-Mail Eligible	NO-MATCH/ NOT EXEMPTION ELIGIBLE/ Not Age Vote-by-Mail Eligible
Anglo	379,424 (4.6%)	301,332 (3.6%)	270,906 (3.3%)
Black	134,988 (7.9%)	126,952 (7.4%)	109,139 (6.4%)
Hispanic	175,769 (5.8%)	151,514 (5.0%)	138,652 (4.6%)
Other	11,515 (2.3%)	9,683 (2.0%)	9,191 (1.9%)
All	701,696 (5.2%)	589,481 (4.4%)	527,888 (3.3%)
	Gross Percentage Point Disparity		
Black% – Anglo%	3.3%	3.8%	3.1%
Hispanic% – Anglo%	1.2%	1.4%	1.3%
	Percent Difference in Rate of NO MATCH		
(Black%-Anglo%) /Anglo%	72%	106%	94%
(Hispanic% - Anglo %) /Anglo%	26%	39%	39%

* Baseline Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.4.A. Percent NO MATCH Among Those Who Voted in 2012*, Using Ecological Regression		
RACE	2010	2012
Anglo	1.0% (+/-0.1)	1.7% (+/-0.04)
Black	4.4% (+/-0.2)	6.2% (+/-0.2)
Hispanic	2.7% (+/- 0.1)	3.1% (+/-0.1)
N	15,652	15,669
R-Squared	.080	.185
	Gross Percentage Point Difference	
Black% – Anglo%	3.4%	4.5%
Hispanic% – Anglo%	1.7%	1.5%
	Relative Rate of NO MATCH	
Black %/Anglo %	340%	265%
Hispanic%/Anglo%	170%	82%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.4.B. Percent NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE Among Those Who Voted in 2010 or 2012*, Using Catalist Racial Data				
	NO MATCH		NO MATCH / NOT EXEMPTION ELIGIBLE	
RACE	2010	2012	2010	2012
Anglo	47,993 (1.4%)	113,507 (2.2%)	42,585 (1.3%)	101,181 (1.9%)
Black	14,954 (2.9%)	45,509 (4.5%)	12,233 (2.4%)	38,194 (3.8%)
Hispanic	13,221 (2.0%)	29,965 (2.2%)	11,032 (1.7%)	25,624 (1.9%)
Other	780 (0.8%)	2,117 (0.9%)	716 (0.7%)	1,947 (0.9%)
All	76,948 (1.7%)	191,098 (2.4%)	66,566 (1.4%)	166,946 (2.1%)
	Gross Percentage Point Disparity			
Black% – Anglo%	1.5%	2.3%	1.1%	1.9%
Hispanic% – Anglo%	0.6%	0.0%	0.4%	0.0%
	Percent Difference in Rate of NO MATCH			
(Black%- Anglo%) /Anglo%	107%	105%	85%	100%
(Hispanic% - Anglo %) /Anglo%	43%	0%	31%	0%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.1.A. Rates of NO MATCHES by Race Under Varying Definitions of the Potential Pool of Registered Voters* Ecological Regressions Using Census Racial Data				
Race	Excluding Catalist Deceased, Deadwood, or NCOA	Excluding Suspense Voters From Pool of Registered	Excluding Expired ID from Pool of Registered	Excluding Catalist Flagged Records or Suspense Voter or Expired ID
Anglo	3.4% (+/- 0.1)	2.9% (+/- 0.1)	2.1% (+/- 0.05)	1.9% (+/- 0.05)
Black	11.3% (+/-0.2)	10.2% (+/ 0.2)	7.6% (+/- 0.2)	7.3% (+/- 0.2)
Hispanic	8.5% (+/- 0.2)	7.9% (+/- 0.2)	5.4% (+/- 0.1)	5.2% (+/- 0.1)
N	15,672	15,670	15,673	15,672
R-Squared	.358	.328	.276	.270
	Gross Percentage Point Disparity			
Black% – Anglo%	8.2%	7.9%	5.5%	5.4%
Hispanic% – Anglo%	5.1%	5.0%	3.3%	3.3%
	Percent Difference in Rate of NO MATCH			
(Black%- Anglo%) /Anglo%	232%	252%	262%	282%
(Hispanic% - Anglo %) /Anglo%	150%	172%	157%	174%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

98

Table VII.1.B. Rates of NO MATCHES by Race Under Varying Definitions of the Potential Pool of Registered Voters* Using Catalist Racial Data				
Race	Excluding Catalist Deceased, Deadwood, or NCOA	Excluding Suspense Voters From Pool of Registered	Excluding Expired ID from Pool of Registered	Excluding Catalist Flagged Records or Suspense Voter or Expired ID
Anglo	364,706 (4.9%)	315,418 (4.3%)	245,510 (3.6%)	167,712 (2.5%)
Black	137,789 (9.0%)	123,964 (8.3%)	95,917 (5.9%)	86,591 (5.1%)
Hispanic	174,952 (6.3%)	159,307 (5.8%)	115,488 (4.0%)	86,591 (3.5%)
Others	10,940 (2.4%)	9,367 (2.1%)	8,088 (1.7%)	5,750 (1.4%)
All	688,387 (5.6%)	608,056 (5.1%)	465,003 (3.6%)	327,132 (3.0%)
	Gross Percentage Point Disparity			
Black% – Anglo%	4.1%	4.0%	2.3%	2.6%
Hispanic% – Anglo%	1.4%	1.5%	0.4%	1.0%
	Percent Difference in Rate of NO MATCH			
(Black%-Anglo%) /Anglo%	84%	93%	64%	104%
(Hispanic% - Anglo %) /Anglo%	29%	35%	11%	40%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Race	NO MATCH	MATCH
Table VII.2. Validation of Results With Alternative Racial Classification: Using Records With the Highest Confidence in the Racial Classification*		
Anglo	223,113 (4.6%)	4,591,961 (95.4%)
Black	51,529 (10.7%)	428,483 (89.3%)
Hispanic	77,370 (6.2%)	1,179,969 (92.9%)
Other	4,252 (1.9%)	216,250 (98.1%)
All	356,264 (5.3%)	6,416,663 (94.7%)
	Gross Percentage Point Disparity	
Black% – Anglo%	6.1%	
Hispanic% – Anglo%	1.6%	
	Percent Difference in Rate of NO MATCH	
(Black%-Anglo%) /Anglo%	133%	
(Hispanic% - Anglo %) /Anglo%	35%	

*Universe: All Registration Records in TEAM Assign a Racial Classification with High Confidence less records indicated as Deceased by State of Texas Database

Table VII.3. Validation of Results With Alternative Racial Classification Using Spanish Surname Voter Registrations: Comparison of No-Match rates of Spanish Surname Registered Voters and Others*		
Race	NO MATCH	MATCH
SSVR	195,398 (6.4%)	2,878,228 (93.6%)
Non-SSVR	591,321 (5.7%)	9,822,638 (93.6%)
All	786,719 (5.8%)	12,700,866 (94.2%)
	Gross Percentage Point Disparity	
SSVR – Non-SSVR	0.7%	
	Percent Difference in Rate of NO MATCH	
(SSVR – Non-SSVR)/Non-SSVR	12%	

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.4. Identification Match versus Disability Exemption Eligible*				
		Disability Exemption Eligible		
STATE OR FEDERAL ID		No Match to Disability	Match to Disability	ALL
	No Match	701,696	85,031	786,727
	Match	11,839,587	861,280	12,700,867
	ALL	12,541,283	946,311	13,487,594

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VIII.1a. Percent of Registered Anglos, Hispanics, and Blacks in Catalist Database who Voted in the State of Texas in 2010 and 2012: Current Active and Suspense Voters						
	2010			2012		
	% Voted	Number Voted	Number Not Voting	% Voted	Number Voted	Number Not Voting
Anglo	41.8%	3,364,053	4,689,493	64.3%	5,169,740	2,874,078
Hispanic	22.0%	655,046	2,320,565	45.0%	1,340,119	1,635,492
Black	31.3%	527,216	1,156,753	59.8%	1,007,153	676,821
	Gross Percentage Point Disparity					
Anglo% Vote-Black%Vote	10.5%			4.5%		
Anglo% Vote-Hisp% Vote	19.8%			20.3%		
	Percent Difference in Rate of Voting					
(Black%-Anglo%)/Anglo%	34%			8%		
(Hispanic% -Anglo %)/Anglo%	90%			43%		

103

Table VIII.1b. Percent of Registered Anglos, Hispanics, and Blacks in Catalist Database who Voted in the State of Texas in 2010 and 2012: Current Active Voters Only						
	2010			2012		
	% Voted	Number Voted	Number Not Vote	% Voted	Number Voted	Number Not Vote
Anglo	44.0%	3,260,374	4,157,144	68.2%	5,056,818	2,360,700
Hispanic	23.1%	636,741	2,121,984	47.6%	1,312,378	1,446,347
Black	33.2%	509,403	1,024,863	63.4%	973,266	561,000
	Gross Percentage Point Disparity					
Anglo % - Black %	9.8%			5.2%		
Anglo % - Hispanic %	20.9%			20.6%		
	Percent Difference in Rate of Voting					
(Black%-Anglo%)/Anglo%	33%			8%		
(Hispanic% - Anglo %)/Anglo%	90%			43%		

104

Table VIII.2a.   Ecological Regression Estimates of Registration as a Percent of Voting Age Population for Anglos, Hispanics, and Blacks in the State of Texas* (95 Percent Confidence Interval in Parentheses)

	2006	2008	2010	2012
Anglo**	82.5%	87.1%	86.0%	84.7%
	(± 1.0)	(± 0.8)	(± 0.8)	(± 0.8)
Hispanic	55.0%	55.0%	53.0%	50.5%
	(± 1.8)	(± 1.6)	(± 1.4)	(± 1.4)
Black	68.5%	70.5%	76.6%	64.7%
	(± 1.6)	(± 2.8)	(± 2.0)	(± 2.6)
Number of Cases (VTDs)	8660	8660	8660	8660
R-Square	.115	.167	.211	.221
	Gross Percentage Point Disparity			
Anglo%-Black%	14.0%	16.6%	9.4%	20.0%
Anglo%-Hispanic%	27.5%	32.1%	33.0%	34.2%
	Percent Difference in Rate of Registration			
(Black%-Anglo%)/Anglo%	20%	24%	12%	31%
(Hispanic% - Anglo %)/Anglo%	50%	58%	62%	68%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.

** Includes Others.

105

Table VIII.2b. Ecological Regression Estimates of Registration as a Percent of Citizen Voting Age Population for Anglos, Hispanics, and Blacks in the State of Texas*(95 percent confidence intervals in parentheses)				
	2006	2008	2010	2012
Anglo**	83.6% (± 0.9)	88.0% (± 0.8)	86.8% (± 0.7)	87.2% (± 0.7)
Hispanic	80.7% (± 2.1)	81.4% (± 1.9)	79.0% (± 1.4)	75.7% (± 1.4)
Black	80.1% (± 3.0)	81.6% (± 2.6)	77.0% (± 2.0)	74.7% (± 2.5)
Number of Cases (VTDs)	8655	8655	8655	8660
R-Square	.001	.007	.007	.026
Gross Percentage Point Disparity				
Anglo%- Black%	3.4%	6.4%	9.8%	12.6%
Anglo%- Hispanic%	2.9%	6.6%	7.8%	11.6%
Percent Difference in Rate of Registration				
(Black%- Anglo%) /Anglo%	4%	8%	12%	17%
(Hispanic% - Anglo %) /Anglo%	4%	8%	10%	15%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.
** Includes Others.

Table VIII.3a. Ecological Regression Estimates of Voting Rates Among Groups as a Percent of Voting Age Population of Anglos, Hispanics, and Blacks in the State of Texas* (95 Percent Confidence Interval in Parentheses)				
	2006	2008	2010	2012
Anglo**	36.4% (± 0.4)	60.8% (± 0.6)	40.6% (± 0.4)	61.1% (± 0.5)
Hispanic	8.4% (± 0.7)	19.2% (± 0.6)	8.5% (± 0.8)	18.1% (± 1.0)
Black	12.8% (± 1.2)	40.2% (± 1.9)	19.4% (± 1.3)	39.2% (± 1.8)
Number of Cases (VTDs)	8660	8660	8660	8660
R-Square	.398	.392	.442	.431
	Gross Percentage Point Disparity			
Anglo%-Black%	23.6%	20.6%	21.2%	21.8%
Anglo%-Hispanic%	28.0%	41.6%	32.2%	43.0%
	Percent Difference in Rate of Voting			
(Black%-Anglo%) /Anglo%	184%	51%	108%	56%
(Hispanic% - Anglo %) /Anglo%	333%	217%	378%	238%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.

** Includes Others.

107

Table VIII.3b. Ecological Regression Estimates of Voting Rates Among Groups as a Percent of Citizen Voting Age Population of Anglos, Hispanics, and Blacks in the State of Texas* (95 Percent Confidence Interval in Parentheses)				
	2006	2008	2010	2012
Anglo**	36.8% (± 0.4)	61.6% (± 0.5)	41.1% (± 0.4)	61.8% (± 0.5)
Hispanic	13.2% (± 0.9)	29.9% (± 0.6)	13.6% (± 0.9)	28.8% (± 1.2)
Black	15.0% (± 1.2)	44.7% (± 1.9)	21.8% (± 1.3)	43.3% (± 1.8)
Number of Cases (VTDs)	8655	8655	8655	8655
R-Square	.273	.212	.296	.244
	Gross Percentage Point Disparity			
Anglo%-Black%	21.8%	16.9%	19.3%	18.4%
Anglo%-Hispanic%	23.6%	31.7%	27.4%	33.1%
	Percent Difference in Rate of Voting			
(Black%-Anglo%) /Anglo%	145%	38%	89%	43%
(Hispanic% - Anglo %) /Anglo%	179%	106%	202%	115%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.

** Includes Others.

108

Table VIII.4. Current Population Survey Estimates of Percent of Anglo, Hispanic, and Black Adult Citizens who are Registered and who Voted in the State of Texas*				
	Percent Reported Being Registered**			
	2006	2008	2010	2012
Anglo	72.8% (± 3.2)	73.6% (± 3.0)	66.9% (± 3.2)	73.0 % (± 3.0)
Hispanic	58.1% (± 7.0)	54.3% (± 6.0)	53.3% (± 7.2)	54.5% (± 6.8)
Black	64.6% (± 8.8)	74.0% (± 7.2)	61.2% (± 7.8)	72.8 % (± 7.0)
	Percent Reported Voting**			
	2006	2008	2010	2012
Anglo	45.2% (± 3.6)	64.7% (± 3.2)	43.8% (± 3.2)	60.9 % (± 3.2)
Hispanic	25.4% (± 6.2)	37.8% (± 5.8)	23.1% (± 6.0)	38.8% (± 6.6)
Black	36.6% (± 9.0)	65.8% (± 7.8)	37.7% (± 7.8)	62.5% (± 7.6)

*Source: Current Population Survey, various years, "Voting and Registration and Supplement," Table 4b. http://www.census.gov/hhes/www/socdemo/ voting/publications/p20/index.htm (last accessed June 6, 2014).
**95 Percent Confidence Interval in Parentheses.

Table VIII.5. Current Population Survey Estimates of Percent of Registered Anglos, Hispanics, and Blacks who Voted in the State of Texas from 2006 to 2012*				
	2006	2008	2010	2012
Anglo	62.5% (± 3.2)	87.8% (± 3.0)	65.4% (± 3.0)	83.4 % (± 3.0)
Hispanic	43.8% (± 7.0)	69.5% (± 6.0)	43.4% (± 7.2)	71.3% (± 6.8)
Black	56.7% (± 8.8)	88.5% (± 7.6)	61.7% (± 7.8)	85.9% (± 7.0)
	Gross Percentage Point Disparity			
Anglo% - Black%	5.8%	-0.7%	3.7%	-2.5%
Anglo% - Hispanic%	18.7%	18.3%	22.0%	12.1%
	Percent Difference in Rate of Voting			
(Black%- Anglo%) /Anglo%	10%	-1%	06%	-3%
(Hispanic% - Anglo %) /Anglo%	43%	26%	51%	17%

*Source: Current Population Survey, various years, "Voting and Registration and Supplement," Table 4b. http://www.census.gov/hhes/www/socdemo/voting/ publications/p20/index.htm(last accessed June 6, 2014).

Table X.1. Comparison of SSN9 Match and A/D/G/N Match Rates			
Primary Match Using A/D/G/N Combinations	SSN MATCH		
	NO SSN MATCH	SSN MATCH	Total
NO A/D/G/N MATCH	1,545,754	131,885	1,677,639
A/D/G/N MATCH	118,902	4,909,281	5,028,183
Total	1,664,656	5,041,166	6,705,822

111

APPENDIX

List of Documents Appended Following Appendix Tables

Ex. A – DOJ Algorithm

Ex. B – Texas Algorithm

Ex. C – U.S. State Department Declaration

Ex. D – U.S. Defense Department Declaration

Ex. E – U.S. Citizenship and Immigration Services Declaration

Ex. F– U.S. Social Security Administration Declaration

Ex. G – U.S. Department of Veterans Affairs, Veterans Benefits Administration Declaration

Ex. H – U.S. Department of Veterans Affairs, Veterans Health Administration Declaration

Ex. I – Rebuttal Report of Professor Charles Stewart, *South Carolina v. United States* (D.D.C.).

Ex. J – Curriculum Vitae of Dr. Stephen Ansolabehere

Table A.V.1. Completeness and Uniqueness of Identifiers on TEAM and State of Texas DPS Driver License Databases					
Identifier		Texas Registered Voter List (TEAM)		Texas Department of Public Safety (DPS) (DL only)	
		Percent Complete	Percent Unique	Percent Complete	Percent Unique
Primary Matches	A	99.7	100.0	100.0	95.7
	B	99.8	99.5	100.0	95.5
	C	99.8	99.2	100.0	95.4
	D	99.9	99.8	100.0	95.7
	E	99.8	98.0	100.0	94.1
	F	99.8	97.9	100.0	95.4
	M	76.3	98.7	100.0	95.7
Secondary Matches	G	85.0	98.3	88.1	95.9
	H	62.3	99.7	100.0	95.7
	I	62.3	88.7	100.0	95.7
	K	85.0	98.3	88.1	95.9
	L	0.7	100.0	1.1	88.4
	SSN	49.5	98.8	100.0	95.6
Number of Records		13,564,416		16,052,332	

Table A.V.2. Completeness and Uniqueness of Identifiers on State of Texas Voter DPS Public Safety (ID) and Texas License to Carry Databases					
Identifier		DPS License to Carry		DPS Personal Identification Card	
		Percent Complete	Percent Unique	Percent Complete	Percent Unique
Primary Matches	A	99.9	100.0	100.0	95.6
	B	99.9	99.5	100.0	96.3
	C	99.9	99.2	100.0	96.2
	D	99.9	99.8	100.0	96.0
	E	100.0	98.0	100.0	96.1
	F	100.0	97.9	100.0	96.4
	M	98.0	98.7	100.0	96.7
Secondary Matches	G	87.6	98.3	79.2	96.6
	H	99.9	99.7	100.0	95.7
	I	99.9	98.0	100.0	95.7
	K	87.6	98.0	79.2	95.9
	L	0.01	98.6	1.7	88.4
	SSN	99.9	98.2	99.5	95.6
Number of Records		835,536		3,396,657	

114

Table A.V3. Number Matches of TEAM records to State of Texas Identification Databases By Record Identifier				
Identifier		DPS Driver License	DPS Public Safety ID	DPS License to Carry
Primary Matches	A	8,247,986	787,346	533,958
	B	8,422,296	805,092	545,814
	C	8,557,902	829,268	552,372
	D	8,281,572	790,781	535,897
	E	8,195,758	802,261	528,436
	F	9,763,774	1,107,209	676,170
	M	7,953,620	504,694	290,779
Secondary Matches	G	61,458	14,209	5,316
	H	27,408	9,736	2,189
	I	43,291	9,367	3,902
	K	64,523	15,690	5,501
	L	608	226	50
	SSN	107,766	27,451	10,750
Any Identifier		10,663,738	1,284,658	733,008

115

Table A.V4. Number Matches of TEAM records to Federal Identification Databases By Record Identifier				
Identifier		DOD Military Identification	DOS Passport or Passport Card	USCIS Citizenship Document
Primary Matches	A	350,515	3,018,662	207,387
	B	357,465	3,107,218	221,813
	C	365,985	3,202,237	213,238
	D	352,008	3,036,730	236,996
	E	348,402	3,000,346	206,663
	F	433,861	4,521,683	304,530
Secondary Matches	G	786	11,404	30,225
	H	1,569	31,747	51,524
	I	2,379	23,477	60,063
	K	1,111	16,810	17,378
	L	163	1,001	1,817
	SSN	8,127	150,323	52,938
Any Nationwide Match		166,128	5,386,264	323,425
Any Identifier		638,354	5,731,163	735,086

116

Table A.V5. Number Matches of TEAM records to Federal Identification and Exemption Databases By Record Identifier				
Identifier		VA (VHA) VIC/VHIC (Identification)	VA (VBA) Disability (Exemption)	SSA Disability (Exemption)
Primary Matches	A	180,483	100,758	412,295
	B	186,754	103,528	434,720
	C	190,053	105,868	450,148
	D	181,146	101,347	417,348
	E	181,332	101,491	420,879
	F	235,188	146,009	636,806
Secondary Matches	G	454	120,591	2,238
	H	1,063	78,406	9,115
	I	1,509	95,516	5,096
	K	617	120,918	2,704
	L	38	391	236
	SSN	5,951	78,191	268,58
Any Nationwide Match		33,487	25,228	73,236
Any Identifier		296,005	188,516	804,338

117

Table A.VI.1. Incidence of NO-MATCH By Racial Group, Using Catalist Racial Classification* and Excluding Records indicated as NCOA, Deceased or Deadwood by Catalist

Race	No Match / Not EXEMPTION ELIGIBLE	No Match / Not Age Vote-By-Mail Eligible	No Match / Not EXEMPTION ELIGIBLE/ Not Age Vote-by-Mail Eligible
Anglo	326,153 (4.4%)	262,330 (3.5%)	235,216 (3.2%)
Black	117,322 (7.7%)	138,928 (7.3%)	95,984 (6.3%)
Hispanic	158,199 (5.7%)	138,776 (5.0%)	127,077 (4.6%)
Other	10,316 (2.3%)	8,714 (1.9%)	8,277 (1.8%)
All	611,990 (5.0%)	521,748 (4.3%)	466,554 (3.8%)
	Gross Percentage Point Disparities		
Black % - Anglo %	3.3%	3.8%	3.1%
Hispanic % – Anglo %	1.3%	1.5%	1.4%
	Relative Rate of NO MATCH		
Black % / Anglo %	1.75	2.09	1.97
Hispanic % / Anglo %	1.30	1.43	1.44

118

Exhibit A

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,
 	
             Plaintiffs,
 	
      v.
 	
RICK PERRY, *et al.*,
 	
             Defendants.
	Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,
 	
             Plaintiff,
 	
TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,
 	
             Plaintiff-Intervenors,
 	
TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,
 	
             Plaintiff-Intervenors,
 	
      v.
 	
STATE OF TEXAS, *et al.*,
 	
             Defendants.
	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

                    Plaintiffs,

          v.

NANDITA BERRY, *et al.*,

                    Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

                    Plaintiffs,

          v.

STATE OF TEXAS, *et al.*,

                    Defendants

Civil Action No. 2:13-cv-348 (NGR)

United States' Database Matching Protocol

The United States hereby sets out its protocol for comparisons among and between the

Texas databases and the federal databases. The matching process proceeds in four parts, which

are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier

values are constructed by combining multiple individual fields. *Third*, one-to-many matches are

conducted between databases. *Fourth*, Texas data concerning deceased individuals is appended

to matching output. Each of the four parts is divided into multiple stages, which in turn are

divided into concrete steps.

The database preparation and standardization steps below apply to both the matches

requested by the United States and all Plaintiffs, as well as the matches requested by Defendants.

1

GLOSSARY

The following definitions apply to the terms used in this document.

- **Disability database:** Any federal database containing records that indicate whether an individual has a disability that would permit that individual apply for an exemption from the SB 14 identification requirement.

- **Identification database:** Any Texas or federal database containing records that indicate whether an individual has identification required by SB 14 in most cases to cast an in-person ballot

- **Protocol:** The instructions to prepare data for matching and to conduct multiple matching sweeps between databases according to specified search criteria.

- **Sweep:** A comparison of a set of database fields in the TEAM database against a set of database fields in a disability database or an identification database.

- **TEAM database:** The Texas Election Administration Management database, the state database that contains records of registered voters.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from TEAM, Identification & Disability Databases

Step 1.1.1:   Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2:   Extract date of birth.

Step 1.1.3:   Extract gender.

Step 1.1.4:   Extract residential address and mailing address.

Step 1.1.5:   Extract social security number.

Step 1.1.6:   Extract Texas driver license number (only if present in Federal database).

Step 1.1.7:   Extract unique record identifier (such as VUID in the TEAM database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1:   Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2:   Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Step 1.2.3:   Separate the identification databases produced by the State of Texas into separate files for each form of state identification at issue, namely driver licenses, identification cards, licenses to carry concealed handguns, and election identification certificates.

3

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Stage 5: Standardize First Name and Middle Name

Step 1.5.1:   Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2:   Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3:   Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1:   Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2:   Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1:   Code gender as a string of 1 for females and 0 for males.

Step 1.7.2:   Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3:   If missing values remain, code all missing values as blank fields.

---

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1:   Convert the residential ZIP code to a string if it is stored as a numeric

field.

Step 1.8.2:   Where the residential address ZIP code is blank, populate that field with

the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3:   Truncate the residential ZIP code field to the first five digits. *E.g.,*

<77777-1234> becomes <77777>.

Step 1.8.4:   Where the residential address field is blank, populate that field with the

value in the mailing address field, if available.

Step 1.8.5:   Where address field containing street address begins with a street number,

isolate the street number. *E.g.,* <123 Main Street> becomes <123>.

Step 1.8.6:   Where the address field begins with recognized strings indicating a mail

box, eliminate strings to isolate the box number. *E.g.,* <PO Box 444>

becomes <444>.

Step 1.8.7:   If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1:   Convert the social security number to a string if it is stored as a numeric

field.

Step 1.9.2:   Using full social security number, check for invalid SSNs. In the case of

invalid SSNs, code as missing. *E.g.,* <123456789> becomes <>.

---

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3:   Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4:   Code all missing values as blank fields.

<u>Stage 10: Identical Records</u>

Step 1.10.1:   For the TEAM database, for the small number of records with different VUID but identical first and last name, gender, residential address number and ZIP, date of birth, and SSN, treat these records as identical.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1:   Create Combination A: First Name + Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2:   Create Combination B: Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.3:   Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4:   Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5:   Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6:   Create Combination F: First Name + Last Name + Gender + Date of Birth.

Step 2.1.7:   Create Combination M: Texas Driver License Number (where available).

---

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1:   Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2:   Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3:   Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4:   Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5:   Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6:   Full Social Security Number.

Stage 3: Establish Identifier Uniqueness

Step 2.3.1:   Within the TEAM database, determine the uniqueness of each primary and secondary identifier variable and mark accordingly. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, mark that record as unique for Combination A.   By contrast, if multiple records have the string <JOHNSMITHA0101950> for Combination G, mark all such records as non-unique for Combination G.

Step 2.3.2:   Within the identification and disability databases, generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination, generate a field that establishes

---

[4] Only for the State Department, create three further variations of Combination G using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

9

uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

10

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identification or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identification or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identification databases produced by the State of Texas.

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identification or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[5] in the TEAM database against the equivalent combination/field in the identification or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identification or disability database.

---

[5] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

12

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identification or disability database.

Stage 3: United States' Nationwide Federal Database Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identification or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identification or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination

13

K against Combination G, Combination K, and Combination L in the

nationwide identification or disability database.

Step 3.3.4:   For each case in which no matches were found in the primary and

secondary matching sweeps of Texas records in a federal identification or

disability database, use the procedure in Step 3.3.1 to match Combination

L against Combination G, Combination K, and Combination L in the

nationwide identification or disability database.[6]

---

[6] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:
- Combination G1 to DOB + Last fix + First fix + Middle name from the TEAM database;
- Combination G2 to DOB + Last fix + First fix from the TEAM database;
- Combination G2 to DOB + Last fix + First word of First name from the TEAM database;
- Combination G3 to DOB + Last fix + First fix + Middle Initial from the TEAM database;
- Combination G3 to DOB + Last fix + First word of First name + Middle Initial from the TEAM database; and
- Combination G3 to DOB + Last fix + First word of First name + First character of Second word of First fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

PART IV: DATA GATHERING

Stage 1: Gather Information Regarding Deceased Individuals

Step 4.1.1:   In all instances in which a unique match was achieved between the TEAM database and the Texas Driver License, Texas Identification Card, and Texas Concealed Handgun database, determine whether the identification record at issue has been flagged as deceased in the Texas identification database.

Step 4.1.2:   Where the driver license, concealed handgun license, or Texas identification card record at issue has been flagged as deceased, append the "deceased" flag to the TEAM record to which the unique match has been made.

Criteria for Matching Records in the TEAM Database

ID Matching			
Database	**Valid ID**	**Matching Algorithm(s)**	**Additional Parameters**
Texas Dept. of Public Safety	Driver's License	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Personal Identification Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Concealed Handgun License	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Texas Election Identification Certificate	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
U.S. Department of Defense	Military ID-Common Access Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Military ID- Uniformed Services ID Cards	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Military ID- DOD Civilian Retiree Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted

U.S. Dept. of Veterans Affairs	Veterans Affairs ID Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None.
U.S. Dept. of Homeland Security	U.S. Certificate of Citizenship	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None
	U.S. Certificate of Naturalization	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None
U.S. State Department	U.S. Passport	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	U.S. Passport Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
Permanent Exemptions			
U.S. Social Security Administration	N/A	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	Determined to have a disability
U.S. Department of Veterans Affairs	N/A	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	Disability rating of 50% or more

Notes:

For each ID/Database combination outlined above all specified matching queries should be run. Based on the matching algorithms, if a record (registrant) in the TEAM database matches a record in the database being searched, then this will be denoted as a match (For any given type of SB14 ID, a registrant will be determined to have valid ID if matched by any of the matching algorithms outlined). New fields will be added to the TEAM database to denote the presence of a match for each database/ID/query. In addition to searching for ID possession, records in the TEAM database will be flagged to indicate the registrant is exempt due to the presence of a qualified disability. (Disclaimer: The matching algorithms here denoted are based on data fields currently known to populate the listed databases).

Explanation of matching queries to be run (numbers correspond to matching algorithms in table above):
1. Records will be matched using the last four digits of the Social Security Number (where available) along with the last name and date of birth fields.
2. Records will be matched using the full Social Security Number field (where available).
3. Records will be matched using the first name, last name, and date of birth fields.
4. Records will be matched using the first name, middle initial, last name, and date of birth fields.
5. Records will be matched using the Texas Driver's License Number field (where available).

Exhibit C

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,
 	
                 Plaintiffs,
 	
        v.
 	
RICK PERRY, *et al.*,
 	
                 Defendants.
	Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,
 	
                 Plaintiff,
 	
TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,
 	
                 Plaintiff-Intervenors,
 	
TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,
 	
                 Plaintiff-Intervenors,
 	
        v.
 	
STATE OF TEXAS, *et al.*,
 	
                 Defendants.
	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

               Plaintiffs,

      v.

JOHN STEEN, *et al.*,

               Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

               Plaintiffs,

      v.

STATE OF TEXAS, *et al.*,

               Defendants

Civil Action No. 2:13-cv-348 (NGR)

CORRECTED DECLARATION OF LEE CHARLES BAYDUSH

I, Lee Charles Baydush, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as subject matter expert for the U.S. Department of State, Bureau of
   Consular Affairs, Consular Systems & Technology, Service Integration & Innovation
   Division (CA/CST/SII). I have served Consular Affairs in this capacity with SII and
   others as a contractor specializing in Passport Systems for more than 28 years. My duties
   include research, analysis, consulting, and operational support of Consular IT systems.

2. I have personal knowledge of the information contained in this declaration based upon
   my work for the State Department, and my personal involvement in completion of the
   database comparison described below.

3. On February 27, 2014, the State Department received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. The State Department was able to load all records except for 6. The identifiers (VUIDs) for 5 of the 6 records that my agency was not able to load and analyze as part of the database comparison process are as follows: 1018504714; 1018510162; 1018555565; 1111894862; 1164253546. The final record that could not be loaded lacked a VUID, but has a unique identifier of 3890395vf (other than this unique identifier "var_fill" value, this record was blank).

4. The Texas data that the State Department received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str

- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- and
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique

-	unique_count	-	FML1D_TAG
-	FMLD	-	FML1D_unique
-	FMLD_TAG	-	FML2D
-	FMLD_unique	-	FML2D_TAG
-	SDZ	-	FML2D_unique
-	SDZ_TAG	-	unique_count2
-	SDZ_unique	-	var_fill
-	SND	-	K2
-	SND_TAG	-	K3
-	SND_unique	-	L2
-	FML1D	-	L3

5. The State Department executed the data preparation and comparison steps, attached as

    Ex. A, that were provided by the United States Department of Justice.

6. Through this process, the State Department attempted to match particular combinations

    of identifying information for Texas registered voters (*e.g.*, first and last name, gender,

    and date of birth) with the same or related combinations of identifying information with

    respect to holders of U.S. Passports and U.S. Passport cards. For each particular

    combination, the State Department identified instances where the identifying information

    for a Texas registered voter matched with the respective combination of identifying

    information for one or more holders of a U.S. Passport or U.S. Passport Card.

7. For each sweep through the relevant State Department datasets, we appended a column to

    the Texas data to indicate, on a record-by-record basis, the output of the database

    comparisons on all of the requested combinations.

8. Two State Department datasets were created for the analysis—one contained qualifying

    records of U.S. Passport and U.S. Passport Card holders with a Texas mailing address

    and the other contained qualifying records of U.S. Passport and U.S. Passport Card

    holders without a Texas mailing address. The frequencies of any missing values in those

    datasets are provided as Ex. B.

9. On April 9, 2014, the State Department transferred responsive data to the Department of Justice, in the form of a variable width file with comma delimiters. In this file, there were 7,741 records in the TEAM database with commas embedded in the address field, and several hundred records with commas embedded in various name fields. Such commas could be incorrectly treated as delimiters in reading the Department of State matching results. As a result, I re-exported the results file using pipe delimiters and on May 14, 2014, the State Department transferred to the Department of Justice a compressed and encrypted 1.7 gigabyte zip file on a DVD that contained a 7.44 gigabyte pipe-delimited file. Both the April 9 and May 14 files are identical other than the use of different delimiters. The use of pipe delimiters on the May 14, 2014, file will prevent any difficulties reading the embedded commas in the results file. The name of the May 14, 2014 results file is Team_Pipe_Delim.

10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that the State Department returned to the Department of Justice on May 14:

   a. "AGDN_matches_id" contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. "AGDNlast_matches_id" contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. "AGD_matches_id" contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. "ADN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. "AGN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. "GDN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination F match;

g. "FMLD_matches_id" contains results of the Stage 2, Step 3.2.1, Combination G match;

h. "SDZ_matches_id" contains results of the Stage 2, Step 3.2.2, Combination H match;

i. "SND_matches_id" contains results of the Stage 2, Step 3.2.2, Combination I match;

j. "ssn_matches_id" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

k. "k2_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

l. "FML1D_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K match;

m. "k3_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "l2_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "FML2D_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. " l3_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "GDN_matches_id_Other_GDN" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "FMLD_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "SND_matches_id_Other_SND" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "ssn_matches_id_Other" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "FML1D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "FML1D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "FML1D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "FML2D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "FML2D_matches_id_Other_FML2D" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "FML2D_matches_id_Other_FML1D" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "a1_matches_id_TX" and "a1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G1 to DOB + Last_fix + First_fix + Middle_name matches for the Texas-only and non-Texas address datasets;

bb. "a2_1_matches_id_TX" and "a2_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First_fix matches for the Texas-only and non-Texas address datasets;

cc. "a2_2_matches_id_TX" and "a2_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First word of First_name matches for the Texas-only and non-Texas address datasets;

dd. "a3_1_matches_id_TX" and "a3_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB + Last_fix + First_fix + Middle_Initial matches for the Texas-only and non-Texas address datasets;

ee. "a3_2_matches_id_TX" and "a3_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + Middle_Initial matches for the Texas-only and non-Texas address datasets;

ff. "a3_3_matches_id_TX" and "a3_3_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + First character of Second word of First_fix matches for the Texas-only and non-Texas address datasets;

gg. "TXMM_snlastdAll_matches_id" contains results of the Stage 4, Step 3.4.1, Sweep 1 match, while "TXMM_TX_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the Texas-only dataset and "TXMM_Other_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the non-Texas addresses dataset;

hh. "TXMM_ssn_matches_id" contains results of the Stage 4, Step 3.4.2, Sweep 2 match, while "TXMM_TX_ssn_matches_id" contains the results for the Sweep 2 match based on the Texas-only dataset and "TXMM_Other_ssn_matches_id" contains the results for the Sweep 2 match based on the non-Texas address dataset;

ii. "TXMM_ndAll_matches_id" contains results of the Stage 4, Step 3.4.3, Sweep 3 match, while "TXMM_TX_ndAll_matches_id" contains the results for the Sweep 3 match based on the Texas-only dataset and "TXMM_Other_ndAll_matches_id" contains the results for the Sweep 3 match based on the non-Texas address dataset; and

jj. "TXMM_fmldAll_matches_id" contains results of the Stage 4, Step 3.4.4, Sweep 4 match, while "TXMM_TX_fmldAll_matches_id" contains the results for the Sweep 4 match based on the Texas-only dataset and "TXMM_Other_fmldAll_matches_id" contains the results for the Sweep 4 match based on the non-Texas address dataset.

11. No other matching comparisons between the Texas data and data from State Department were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on June 24, 2014.

Lu Baydurl

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1:   Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2:   Extract date of birth.

Step 1.1.3:   Extract gender.

Step 1.1.4:   Extract residential address and mailing address.

Step 1.1.5:   Extract social security number.

Step 1.1.6:   Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1:   Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

---

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.,* <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.,* <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.,* <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.,* <123456789> becomes <>.

---

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3:   Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4:   Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1:   Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2:   Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3:   Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4:   Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5:   Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6:   Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7:   Create Combination M: Texas Driver License Number (where available).

---

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1:   Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2:   Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3:   Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4:   Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5:   Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6:   Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1:   Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2:   Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3:   Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

> **Step 2.4.1**:   Generate a field that establishes the uniqueness of each identifier variable.
>
> For federal databases, for each combination A-L, generate a field that
>
> establishes uniqueness among only Texas records and a field that
>
> establishes uniqueness among nationwide records. *E.g.*, if only one record
>
> has the string <JEANSMITH01011950012377777> for Combination A,
>
> populate the uniqueness field for Combination A for that record as <1>. If
>
> four records have the string <JOHNSMITHA0101950> for Combination
>
> G, populate the uniqueness field for Combination G for each of those
>
> records as <2>, which indicates any number greater than one.

---

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1:   For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2:   Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3:   Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

8

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

9

Exhibit A

Step 3.2.4:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

10

Exhibit A

> K against Combination G, Combination K, and Combination L in the
>
> nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and

secondary matching sweeps of Texas records in a federal identifier or

disability database, use the procedure in Step 3.3.1 to match Combination

L against Combination G, Combination K, and Combination L in the

nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

---

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last fix + First fix + Middle name from the TEAM database;
- Combination G2 to DOB + Last fix + First fix from the TEAM database;
- Combination G2 to DOB + Last fix + First word of First name from the TEAM database;
- Combination G3 to DOB + Last fix + First fix + Middle Initial from the TEAM database;
- Combination G3 to DOB + Last fix + First word of First name + Middle Initial from the TEAM database; and
- Combination G3 to DOB + Last fix + First word of First name + First character of Second word of First fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2   Regardless of whether full 9 social security number is unique in the

TEAM database, for Sweep 2, match against the equivalent field in a

nationwide search of the Federal database.

Step 3.4.3   Regardless of whether the combination for Sweep 3 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

Step 3.4.4   Regardless of whether the combination for Sweep 4 (Combination G) is

unique in the TEAM database, match against the equivalent combination

in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:
- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:
- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Step 1.3.1 Diagnostics: Report the Frequency of Missing Values for Each Field

6, 781,544 Qualifying U.S. Passport/U.S. Passport Card records with 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	38
date_of_birth	0
gender_code	5
ssn	25,640
Last_fix	0
multi_lastname1	6,717,750
multi_lastname2	6,717,749
first_fix	38
middle_initial	929,454
fem_str	0
dob_str	0
zip5_num_res_alt	50
ssn4_str	25,640

92,776,332 Qualifying U.S. Passport/Passport Card records without 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	1,596
date_of_birth	2
gender_code	247
ssn	641,147
last_fix	0
multi_lastname1	91,615,428
multi_lastname2	91,615,406
first_fix	1,596
middle_initial	12,567,280
fem_str	16
dob_str	0
zip5_res_alt	757,630
str_num_res_alt	43,961

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

                Plaintiffs,

      v.

RICK PERRY, *et al.*,

           Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

              Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

          Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

         Plaintiff-Intervenors,

      v.

STATE OF TEXAS, *et al.*,

           Defendants.

Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

               Plaintiffs,

       v.

JOHN STEEN, *et al.*,

               Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

               Plaintiffs,

       v.

STATE OF TEXAS, *et al.*,

               Defendants

Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF Michelle Saunders Rudolph

I, Michelle Saunders Rudolph, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Division Director, Data, Analysis and Programs Division, DMDC/DHRA/DoD. I have served in this capacity for close to 2 years. My duties include ensuring the data arriving at DMDC is of sound quality so Servicemembers and their families can receive the benefits and entitlements they have earned. In addition, I ensure agencies and offices with a need to know, receive accurate and timely data.

2. I have personal knowledge of the information contained in this declaration based upon my work for the Defense Manpower Data Center at the U.S. Department of Defense ("DMDC"), and my supervision of completion of the database comparison described below.

3. On February 28, 2014, DMDC received from the United States Department of Justice an

   encrypted hard drive containing data regarding Texas registered voters in the form of a 7

   gigabyte .csv file containing 13,564,420 records for Texas registered voters. DMDC was

   able to load all 13,564,420 records.

4. The Texas data that DMDC received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique

- AGD
- AGD_TAG
- AGD_unique
- and
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique
- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill

- K2
- K3

- L2
- L3

5. DMDC executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

6. Through this process, DMDC attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to persons eighteen years of age or older holding a valid form of one of the following Department of Defense ("DoD") photo identification cards: Common Access Cards and Uniformed Services ID Cards (including Department of Defense Civilian Retiree Cards). For each particular combination, DMDC identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of DoD photo identification.

7. In order to conduct the requested matches, and to create and record the combinations used for the Stage 4 matches, DMDC reformatted the original .csv file into a fixed format flat file to which matching results were appended.

8. Two DoD datasets were created for this analysis—one contained records of holders of qualifying DoD photo ID located in Texas and the other contained records for holders of qualifying DoD photo ID located outside of Texas. For each sweep through the relevant DoD datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of any missing values in the underlying DoD data are provided as Ex. B.

9. On May 9, 2014, DMDC transferred responsive data to the Department of Justice, in the form of a 21.5 gigabyte file with pipe delimiters. The name of the results file transferred to the Department of Justice is "Team Match Report".

10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that DMDC returned to the Department of Justice on May 9:

   a. "Key_A_match_flag_Texas" contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. "Key_B_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. "Key_C_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. "Key_D_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. "Key_E_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. "Key_F_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination F match;

   g. "Key_G_match_flag_Texas" contains results of the Stage 2, Step 3.2.1, Combination G match;

   h. "Key_H_match_flag_Texas" contains results of the Stage 2, Step 3.2.2, Combination H match;

   i. "Key_I_match_flag_Texas" contains results of the Stage 2, Step 3.2.2, Combination I match;

j. "Key_KtoG_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

k. "Key_KtoK_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K match;

l. "Key_KtoL_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

m. "Key_LtoG_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

n. "Key_LtoL_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L match;

o. "Key_LtoK_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

p. "Key_F_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.1, Combination F match;

q. "Key_I_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.2, Combination I match;

r. "Key_KtoG_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

s. "Key_KtoK_match_flag_Nationwide" contains results of the Stage 2, Step 3.3.3, Combination K match;

t. "Key_KtoL_match_flag_Nationwide" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

u. "Key_LtoG_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

v. "Key_LtoL_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L match;

w. "Key_LtoK_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

11. In the May 9 dataset provided to the Department of Justice, no results were provided for the following matches:

a. the Stage 2, Step 3.2.2, nine-digit social security number match;

b. the Stage 3, Step 3.3.2, Combination G match; and

c. the Stage 3, Step 3.3.2, nine-digit social security number match.

These match results were provided separately on to the Department of Justice on May 28, 2014 as described in paragraph 13 below.

12. In the original May 9 dataset provided to the Department of Justice, results for the Stage 4 matches were provided in the following fields:

a. "Sweep_1_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.1, Sweep 1 match;

b. "Sweep_2_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.2, Sweep 2 match;

c. "Sweep_3_match_flag_Nationwide" had partial results for the Sweep 3 match; and

d. "Sweep_4_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.4, Sweep 4 match.

For these Stage 4 matches in the May 9 dataset, the only voter registration records compared were those that had not already matched on a Stage 1 through 3 sweep. Moreover, the only DoD records compared were for qualifying ID holders outside of Texas. At the request of the Department of Justice, DMDC re-ran each of the Stage 4 matches, matching all Texas voter registration records regardless of whether they had been matched on a Stage 1 through 3 sweep. In addition, the complete voter registration database extract was matched against all qualifying DoD records for ID holders in and outside of Texas. These corrected and complete match results for the Stage 4 sweeps were provided to Department of Justice as part of the May 28, 2014 dataset.

13. The May 28 dataset thus supplemented data not previously provided by DMDC as to three of the sweeps requested by the United States, and also re-ran the each of the Stage 4 matches requested by Defendants. To enable the results from the May 28 dataset to be integrated with the May 9 dataset, the May 28 dataset included two unique identifiers: the VUID field from the voter registration file, which is almost completely unique, as well as the Var_Fill field, which is completely unique. In addition to these identifiers, the following results fields were included in the May 28 dataset:

a. "SWEEP_1_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.1, Sweep 1 match based on the Texas-only dataset and "SWEEP_1_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

b. "SWEEP_2_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.2, Sweep 2 match based on the Texas-only dataset and

"SWEEP_2_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

c. "SWEEP_3_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.3, Sweep 3 match based on the Texas-only dataset and

"SWEEP_3_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

d. ""SWEEP_4_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.4, Sweep 4 match based on the Texas-only dataset and

"SWEEP_4_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

e. "SSN9_SECONDARY_TEXAS" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

f. "SSN9_SECONDARY_NATIONWIDE" contains the results of the Stage 3, Step 3.3.2, nine-digit social security number match; and

g. "KEY_G_MATCH_FLAG_NATIONWIDE" contains the results of the Stage 3, Step 3.3.2, Combination G match.

14. In both the May 9 and May 28 datasets, a result of "1" indicates that the respective combination from one DoD record matched; "2" indicates that two or more DoD records matched on the respective combination; "0" indicates that there was no match; and a blank space indicates that no match was attempted.[1]

---

[1] The only exception to this convention is with respect to the Combination G match in the May 28 dataset, where a "0" can indicate both that no match was attempted and that no match was made. No match would have been attempted where the underlying voter registration record had matched on a Stage 1 or Stage 2 combination, or where the particular values in Combination G were not unique among all voter registration records to be compared.

15. No other matching comparisons between the Texas data and data from DMDC were

undertaken beyond those set forth above, and in accordance with the steps provided in

Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

1

Exhibit A

Step 1.2.2:   Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1:   Report the frequency of missing values for each field.

Step 1.3.2:   Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3:   Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1:   Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2:   For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3:   Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4:   Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

 Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

---

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

Step 1.8.1:   Convert the residential ZIP code to a string if it is stored as a numeric

field.

Step 1.8.2:   Where the residential address ZIP code is blank, populate that field with

the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3:   Truncate the residential ZIP code field to the first five digits. *E.g.*,

<77777-1234> becomes <77777>.

Step 1.8.4:   Where the residential address field is blank, populate that field with the

value in the mailing address field, if available.

Step 1.8.5:   Where address field containing street address begins with a street number,

isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6:   Where the address field begins with recognized strings indicating a mail

box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444>

becomes <444>.

Step 1.8.7:   If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1:   Convert the social security number to a string if it is stored as a numeric

field.

Step 1.9.2:   Using full social security number, check for invalid SSNs. In the case of

invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

---

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3:   Extract the last four digits of full social security number as a four-

character string and use them to populate a separate SSN4 field.

Step 1.9.4:   Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1:   Create Combination A: First Name + Last Name + Gender + DOB +

Residential ZIP + Residential Street Number. *E.g.*, the separate fields

<JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are

combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2:   Create Combination B: Last Name + Gender + DOB + Residential ZIP +

Residential Street Number.

Step 2.1.3:   Create Combination C: Gender + Date of Birth + Residential ZIP +

Residential Street Number.

Step 2.1.4:   Create Combination D: First Name + Last Name + Date of Birth +

Residential ZIP + Residential Street Number.

Step 2.1.5:   Create Combination E: First Name + Last Name + Gender + Residential

ZIP + Residential Street Number.

Step 2.1.6:   Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7:   Create Combination M: Texas Driver License Number (where available).

---

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

        Step 2.2.1:   Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

        Step 2.2.2:   Create Combination H: SSN4 + Date of Birth + Residential ZIP.

        Step 2.2.3:   Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

        Step 2.2.4:   Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

        Step 2.2.5:   Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

        Step 2.2.6:   Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

        Step 2.3.1:   Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

        Step 2.3.2:   Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

        Step 2.3.3:   Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

   Step 2.4.1:   Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that

establishes uniqueness among only Texas records and a field that

establishes uniqueness among nationwide records. *E.g.*, if only one record

has the string <JEANSMITH01011950012377777> for Combination A,

populate the uniqueness field for Combination A for that record as <1>. If

four records have the string <JOHNSMITHA0101950> for Combination

G, populate the uniqueness field for Combination G for each of those

records as <2>, which indicates any number greater than one.

---

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

8

Exhibit A

Stage 2: <u>United States' Secondary One-to-Many Matching Sweeps</u>

Step 3.2.1:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

9

Exhibit A

Step 3.2.4:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

> K against Combination G, Combination K, and Combination L in the
>
> nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and

secondary matching sweeps of Texas records in a federal identifier or

disability database, use the procedure in Step 3.3.1 to match Combination

L against Combination G, Combination K, and Combination L in the

nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

---

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last fix + First fix + Middle name from the TEAM database;
- Combination G2 to DOB + Last fix + First fix from the TEAM database;
- Combination G2 to DOB + Last fix + First word of First name from the TEAM database;
- Combination G3 to DOB + Last fix + First fix + Middle Initial from the TEAM database;
- Combination G3 to DOB + Last fix + First word of First name + Middle Initial from the TEAM database; and
- Combination G3 to DOB + Last fix + First word of First name + First character of Second word of First fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2   Regardless of whether full 9 social security number is unique in the

TEAM database, for Sweep 2, match against the equivalent field in a

nationwide search of the Federal database.

Step 3.4.3   Regardless of whether the combination for Sweep 3 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

Step 3.4.4   Regardless of whether the combination for Sweep 4 (Combination G) is

unique in the TEAM database, match against the equivalent combination

in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:
- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:
- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

12

DMDC Step 1.3 Data Diagnostics

Data Element	Data Availability	"In Texas" Population	"Outside of Texas" Population
Last Name	available	746,010	8,285,517
	missing	0	1
First Name	available	746,004	8,285,426
	missing	6	92
Middle Name	available	669,476	7,656,604
	missing	76,534	628,914
SSN	available	746,010	8,285,517
	missing	0	1
Date of Birth	available	746,010	8,285,518
	missing	0	0
Gender	available	746,003	8,285,379
	missing	7	139
Address Text	available	746,010	8,108,935
	missing	0	176,583
Address State	available	746,010	8,026,896
	missing	0	258,622
Address Zip	available	745,993	8,072,943
	missing	17	212,575

Data Element	Data Quality*	"In Texas" Population	"Outside of Texas" Population
SSN	good	744,032	8,211,554
	bad	1,978	73,964
Date of Birth	good	745,936	8,284,603
	bad	74	915

NOTES:

An SSN was considered "bad" if it met any of the following criteria:

*Equaled 000000000, 111111111, 222222222, 333333333,…,999999999

*Equaled sequantial values 123456789 or 234567890

*The first 3 numbers were 000 or 666

*The first number was 9

*The middle node was 00 (ie xxx-00-xxxx)

*The last node was 0000 (ie xxx-xx-0000)

A Date of Birth was considered "bad" if it met either of the following criteria:

*Birth year was 1914 or prior (aka were 100 years of age or older)

*Representative of repeating digits (ie 01/01/01, 02/02/02,…,12/12/12)

Exhibit E

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,
 	
    Plaintiffs,
 	
     v.
 	
RICK PERRY, *et al.*,
 	
    Defendants.
	Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,
 	
    Plaintiff,
 	
TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,
 	
    Plaintiff-Intervenors,
 	
TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,
 	
    Plaintiff-Intervenors,
 	
     v.
 	
STATE OF TEXAS, *et al.*,
 	
    Defendants.
	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

        Plaintiffs,

       v.

JOHN STEEN, *et al.*,

        Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

        Plaintiffs,

       v.

STATE OF TEXAS, *et al.*,

        Defendants

Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF YEMI B. OSHINNAIYE

I, Yemi B. Oshinnaiye, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Associate Chief of Systems Operations, Office of Information

   Technology ("OIT"), U.S. Citizenship and Immigration Services ("USCIS"), Department

   of Homeland Security. I have served in this capacity since May 15, 2014. My duties

   include managing Branch and Project Managers that oversee development, deployment

   and production support for systems that support USCIS current operations. Prior to this

   position I served as Branch Chief for Benefits and Biometrics, USCIS OIT. I served in

   that position beginning in October, 2012. My duties included Management of

   Project/Program managers that oversaw the development, deployment and production

   support for systems that supported the administration, management and granting of

benefits from USCIS and systems that validate eligibility for those benefits. I have personal knowledge of the information contained in this declaration based upon my work for the USCIS, and my personal involvement in and supervision of the database comparison described below.

2. On February 27, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice ("DOJ") in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. USCIS was able to load all 13,564,420 records.

3. The Texas data that USCIS received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt

- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- ADN
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique

- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique

- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill
- K2
- K3
- L2
- L3

4. USCIS executed the data preparation and comparison algorithm as it was relevant to USCIS databases containing U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data , attached as Ex. A, that was provided by DOJ.[1]

5. Through this process, USCIS attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization found in the USCIS Computer Linked Application Information Management System 4 ("CLAIMS 4") and Central Index System ("CIS"), which contain U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data. For each particular combination, USCIS identified instances where the identifying information from the Texas registered voter information matched with the same or related combinations of identifying information for one or more records of individuals issued a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization. The frequencies of any missing values in the underlying USCIS data are provided in Ex. B.

6. For each sweep through the relevant USCIS dataset, we tracked the results,

---

[1] The only data preparation step listed in Ex. A that USCIS was unable to complete was the Step 1.7.2 imputation of gender for USCIS records with addresses outside of Texas. No gender imputation was conducted for USCIS records with an address outside of Texas that contained a blank gender value.
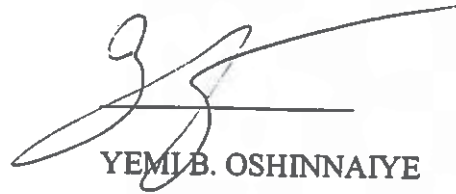
on a record-by-record basis, of the database comparisons on all of the requested combinations. In order to speed computer processing time, the results were appended only to a unique identifier, the Var_Fill value, that had been provided for each Texas voter registration record we received. No other underlying data from the voter registration database was included in the results file sent to DOJ.

7. On May 23, 2014, USCIS transferred responsive data to DOJ, in the form of a 535 megabyte text (.txt) file, with comma delimiters.

8. The results of each sweep, as described in Ex. A, can be found in the following columns of the data USCIS returned to DOJ:

    a. "USA" contains results of the Stage 1, Step 3.1.1, Combination A match;

    b. "USB" contains results of the Stage 1, Step 3.1.2, Combination B match;

    c. "USC" contains results of the Stage 1, Step 3.1.2, Combination C match;

    d. "USD" contains results of the Stage 1, Step 3.1.2, Combination D match;

    e. "USE" contains results of the Stage 1, Step 3.1.2, Combination E match;

    f. "USF" contains results of the Stage 1, Step 3.1.2, Combination F match;

    g. "USG" contains results of the Stage 2, Step 3.2.1, Combination G match;

    h. "USH" contains results of the Stage 2, Step 3.2.2, Combination H match;

    i. "USI" contains results of the Stage 2, Step 3.2.2, Combination I match;

    j. "USSSN" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

    k. "USGK" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

    l. "USK" contains results of the Stage 2, Step 3.2.3, Combination K match;

    m. "USKL" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "USGL" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "USL" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. "USLK" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "USF-NW" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "USG-NW" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "USI-NW" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "USSSN-NW" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "USGK_NW" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "USK_NW" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "USKL_NW" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "USGL_NW" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "USL_NW" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "USLK_NW" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "Texas1" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;

bb. "Texas2" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;

cc. "Texas3" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and

dd. "Texas4" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

9. No other matching comparisons between the Texas data and data from USCIS were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

YEMI B. OSHINNAIYE

Acting Associate Chief of Systems

Operations, Office of Information

Technology, U.S. Citizenship and

Immigration Services

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1:   Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2:   Extract date of birth.

Step 1.1.3:   Extract gender.

Step 1.1.4:   Extract residential address and mailing address.

Step 1.1.5:   Extract social security number.

Step 1.1.6:   Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1:   Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

1

Exhibit A

Step 1.2.2:   Remove records from disability database extracts that do not indicate

current disability status or indicate a Veterans Administration disability

rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1:   Report the frequency of missing values for each field.

Step 1.3.2:   Report the frequencies of invalid Social Security numbers, such as

111111111 and 123456789.

Step 1.3.3:   Report the frequencies of likely invalid dates of birth, such as January 1,

1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1:   Remove last name suffixes that are contained within the last name field,

rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2:   For last names containing hyphens, populate separate last name fields for

all parts of the last name. *E.g.*, the last name <Smith-Jones> would have

the value <Smith> entered into a LastName1 field and the value <Jones>

entered into a LastName2 field.

Step 1.4.3:   Remove spaces, hyphens, periods, and apostrophes from all last name

fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes

<OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4:   Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1:   Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2:   Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3:   Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1:   Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2:   Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1:   Code gender as a string of 1 for females and 0 for males.

Step 1.7.2:   Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3:   If missing values remain, code all missing values as blank fields.

---

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

Step 1.8.1:   Convert the residential ZIP code to a string if it is stored as a numeric

field.

Step 1.8.2:   Where the residential address ZIP code is blank, populate that field with

the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3:   Truncate the residential ZIP code field to the first five digits. *E.g.*,

<77777-1234> becomes <77777>.

Step 1.8.4:   Where the residential address field is blank, populate that field with the

value in the mailing address field, if available.

Step 1.8.5:   Where address field containing street address begins with a street number,

isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6:   Where the address field begins with recognized strings indicating a mail

box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444>

becomes <444>.

Step 1.8.7:   If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1:   Convert the social security number to a string if it is stored as a numeric

field.

Step 1.9.2:   Using full social security number, check for invalid SSNs. In the case of

invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

---

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-

character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

<center>PART II: DATABASE PREPARATION</center>

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB +

Residential ZIP + Residential Street Number. *E.g.*, the separate fields

<JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are

combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP +

Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP +

Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth +

Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential

ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

---

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

<center>5</center>

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1:   Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that

establishes uniqueness among only Texas records and a field that

establishes uniqueness among nationwide records. *E.g.*, if only one record

has the string <JEANSMITH01011950012377777> for Combination A,

populate the uniqueness field for Combination A for that record as <1>. If

four records have the string <JOHNSMITHA0101950> for Combination

G, populate the uniqueness field for Combination G for each of those

records as <2>, which indicates any number greater than one.

---

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1:   For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2:   Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3:   Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

8

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4:   For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3:   For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

10

Exhibit A

K against Combination G, Combination K, and Combination L in the

nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and

secondary matching sweeps of Texas records in a federal identifier or

disability database, use the procedure in Step 3.3.1 to match Combination

L against Combination G, Combination K, and Combination L in the

nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

---

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from
applicable State Department records, first to include only the subset of records with Texas addresses, and
then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas
addresses), against the following fields from the TEAM database:
- Combination G1 to DOB + Last fix + First fix + Middle name from the TEAM database;
- Combination G2 to DOB + Last fix + First fix from the TEAM database;
- Combination G2 to DOB + Last fix + First word of First name from the TEAM database;
- Combination G3 to DOB + Last fix + First fix + Middle Initial from the TEAM database;
- Combination G3 to DOB + Last fix + First word of First name + Middle Initial from the
  TEAM database; and
- Combination G3 to DOB + Last fix + First word of First name + First character of Second word
  of First fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate
<1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2   Regardless of whether full 9 social security number is unique in the

TEAM database, for Sweep 2, match against the equivalent field in a

nationwide search of the Federal database.

Step 3.4.3   Regardless of whether the combination for Sweep 3 is unique in the

TEAM database, match against the equivalent combination in a

nationwide search of the Federal database.

Step 3.4.4   Regardless of whether the combination for Sweep 4 (Combination G) is

unique in the TEAM database, match against the equivalent combination

in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:
- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:
- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

12

Exhibit B

United States Citizenship and Immigration Services

Step 1.3 Diagnostics

USCIS had a population of 19,401,814 United States Citizens who were presented either the N550 Naturalization Certificate or the N560 Certificate of Citizenship. This listing was extracted from the USCIS Central Index System which is the most comprehensive listing of United States Citizens.

Below are the numbers of USCIS records analyzed missing information from a particular listed field.

Step 1.3.1

Missing first name – 49,814
Missing middle name – 7,180,037
Missing last name – 0
Missing date of birth – 18,632
Missing gender – 8,251,075
Missing SSN – 14,264,377

Step 1.3.2:

Number of Records with Invalid SSNs:

5	123456789
2	987654321
12	111111111
3	555555555
2	666666666
1	777777777
2	888888888
40	999999999

Step 1.3.3

Number of Records with Potentially Invalid dates of birth

85	19010101
49	19020202
29	19030303
41	19040404
59	19050505
44	19060606
55	19070707
58	19080808
69	19090909
103	19111111
122	19121212

Exhibit F

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,
 	
          Plaintiffs,
 	
      v.
 	
RICK PERRY, *et al.*,
 	
          Defendants.
	Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,
 	
          Plaintiff,
 	
TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*,
 	
          Plaintiff-Intervenors,
 	
TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*,
 	
          Plaintiff-Intervenors,
 	
      v.
 	
STATE OF TEXAS, *et al.*,
 	
          Defendants.
	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF STEVE STRAUSLER

I, Steve Strausler, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Computer Specialist with the Social Security Administration ("SSA"). I have served in this capacity for 33 years. My duties include updating and maintenance of SSA's Master Beneficiary Record, System of Records Number 60-0090 (published at 71 F.R. 1826 (January 11, 2006)).

2. I have personal knowledge of the information contained in this declaration based upon my work for SSA, and my personal involvement in completion of the database comparison described below.

2

3. On February 28, 2014, SSA received from the United States Department of Justice

    ("DOJ") an encrypted hard drive containing data regarding Texas registered voters in the

    form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters.

    SSA was unable to use this variable width file. We requested that the DOJ provide a

    fixed width file, which we received from DOJ on April 3, 2014. 13,564,419 records were

    provided in the fixed width file and SSA was able to load all of them.[1] In order to hold

    down the file size and make the fixed width file more manageable for SSA to load, this

    file contained only the fields that were strictly necessary to conduct the database

    comparison process, rather than all fields that were used to construct the matching

    combinations, as had been sent in the initial variable width file.

4. The Texas data on the 9.4 GB fixed width file that SSA received contained the following

    fields:

---

[1] One record from the variable width file was omitted because of data formatting issues with that record.

- vuid
- ssn
- ssn4_str
- zip_str
- last_fix
- first_fix
- middle_initial
- dob_str
- AGDN
- AGDNlast
- AGD
- ADN
- AGN

- GDN
- DLN
- FMLD
- SDZ
- SND
- FMLiD
- FML2D
- var_fill
- K2
- K3
- L2
- L3

5. SSA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the DOJ.

6. Through this process, SSA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to records from SSA's master beneficiary records[2] of persons who are 18 years of age or older, and who SSA has determined to be in current payment status for disability.[3]

7. For each sweep through the relevant SSA data, we appended results to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of missing values in the underlying SSA data are provided as Ex. B.

---

[2] SSA's Master Beneficiary Records includes records from the following SSA systems of records: (1) Master Beneficiary Record, System of Record Notice 60-0090 (last published at 71 F.R. 1826 (January 11, 2006)) and (2) Supplemental Security Income Record and Special Veterans Benefits, System of Record Notice 60-0103 (last published at 71 F.R. 1830 (January 11, 2006)).

[3] For Supplemental Security Income (SSI) purposes, SSA's search included individuals who are in active disability status, in addition to individuals who have been determined to be blind.

4

8. On May 28, 2014, SSA transferred responsive data to the Department of Justice, in the form of a 9.75 gigabyte, fixed width, encrypted text file. The name of the results file transferred to the Department of Justice is texas.dojfinal.0527.[4]

9. The results of each sweep, as described in Ex. A, can be found in the following fixed positions in the data that SSA returned to the Department of Justice on May 28:

   a. Position 741 contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. Position 742 contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. Position 743 contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. Position 744 contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. Position 745 contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. Position 746 contains results of the Stage 1, Step 3.1.2, Combination F match;

   g. Position 747 contains results of the Stage 2, Step 3.2.1, Combination G match;

   h. Position 748 contains results of the Stage 2, Step 3.2.2, Combination H match;

   i. Position 749 contains results of the Stage 2, Step 3.2.2, Combination I match;

   j. Position 756 contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

   k. Position 750 contains results of the Stage 2, Step 3.2.3, Combination K to G match;

   l. Position 751 contains results of the Stage 2, Step 3.2.3, Combination K match;

   m. Position 752 contains results of the Stage 2, Step 3.2.3, Combination K to L match;

   n. Position 753 contains results of the Stage 2, Step 3.2.4, Combination L to G match;

   o. Position 755 contains results of the Stage 2, Step 3.2.4, Combination L match;

---

[4] SSA transferred identical results files to the Department of Justice on May 16, 2014 and May 23, 2014. These files could not be accessed by Department of Justice staff because of problems unencrypting the file.

5